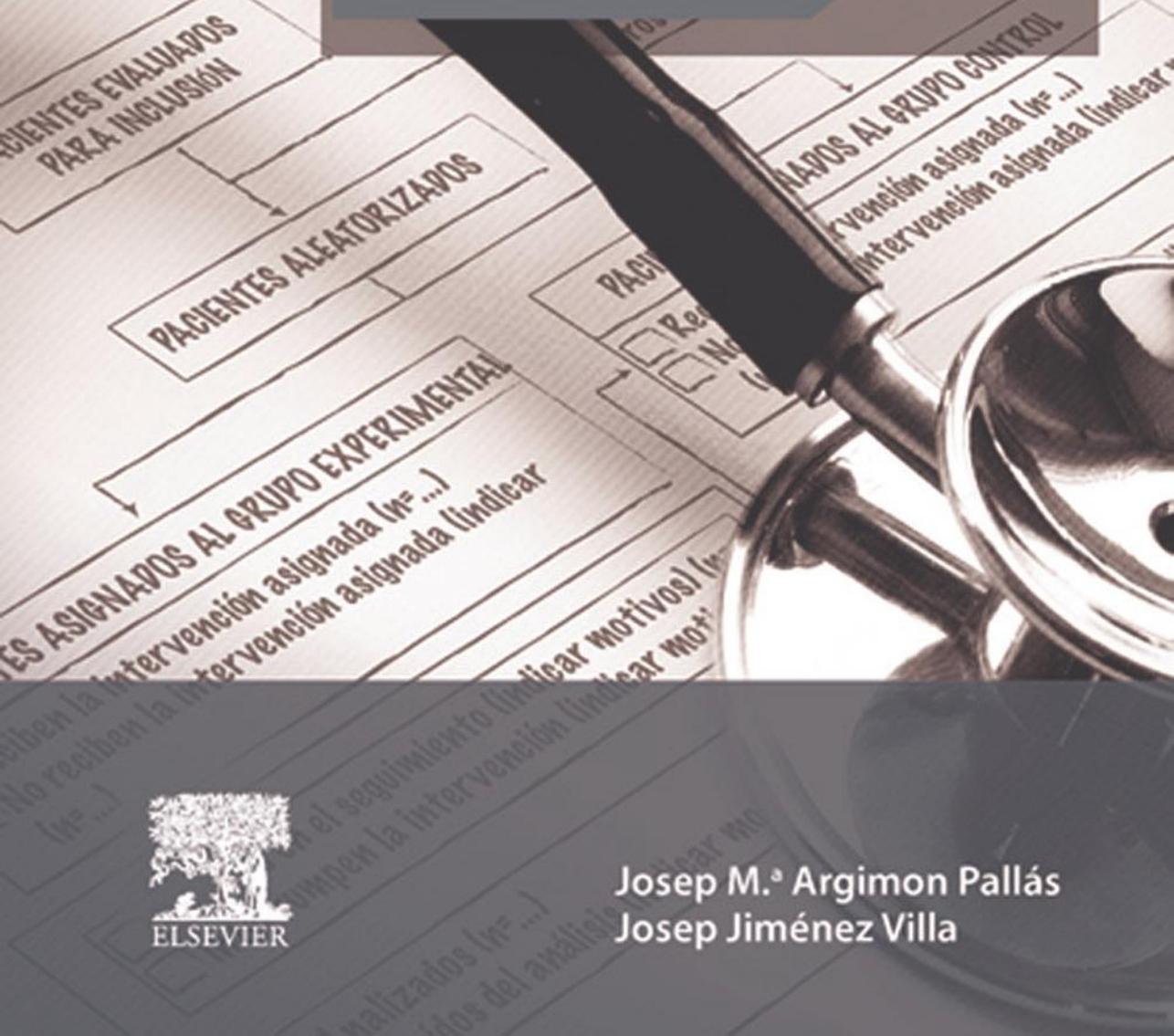




# MÉTODOS DE INVESTIGACIÓN CLÍNICA Y EPIDEMIOLÓGICA

4.º EDICIÓN



Josep M.ª Argimon Pallás  
Josep Jiménez Villa

# **Métodos de investigación clínica y epidemiológica**

Página deliberadamente en blanco

# Métodos de investigación clínica y epidemiológica

4.<sup>a</sup> edición

## **Josep M.<sup>a</sup> Argimon Pallás**

Doctor en Medicina y Cirugía

Especialista en Medicina Preventiva y Salud Pública

Master of Science (M. Sc) en Epidemiología

Master of Science (M. Sc) en Atención Sanitaria basada en la evidencia

Servicio Catalán de la Salud

## **Josep Jiménez Villa**

Doctor en Medicina y Cirugía

Especialista en Medicina de Familia y Comunitaria

División de Evaluación de Servicios Sanitarios. Servicio Catalán de la Salud



Ámsterdam Barcelona Beijing Boston Filadelfia Londres Madrid  
México Milán Múnich Orlando París Roma Sidney Tokio Toronto



ELSEVIER

Primera edición 1991  
Segunda edición 2000  
Tercera edición 2009  
Cuarta edición 2013

© 2013 Elsevier España, S.L.  
Travessera de Gràcia, 17-21 - 08021 Barcelona, España

**Fotocopiar es un delito (Art. 270 C.P.)**

Para que existan libros es necesario el trabajo de un importante colectivo (autores, traductores, dibujantes, correctores, impresores, editores...). El principal beneficiario de ese esfuerzo es el lector que aprovecha su contenido.

Quien fotocopia un libro, en las circunstancias previstas por la ley, delinque y contribuye a la «no» existencia de nuevas ediciones. Además, a corto plazo, encarece el precio de las ya existentes.

Este libro está legalmente protegido por los derechos de propiedad intelectual. Cualquier uso fuera de los límites establecidos por la legislación vigente, sin el consentimiento del editor, es ilegal. Esto se aplica en particular a la reproducción, fotocopia, traducción, grabación o cualquier otro sistema de recuperación de almacenaje de información.

ISBN: 978-84-8086-941-6

Depósito Legal: B. 21.497-2012

Coordinación y producción editorial: Fotolettra S.A.

**Advertencia**

La medicina es un área en constante evolución. Aunque deben seguirse unas precauciones de seguridad estándar, a medida que aumenten nuestros conocimientos gracias a la investigación básica y clínica habrá que introducir cambios en los tratamientos y en los fármacos. En consecuencia, se recomienda a los lectores que analicen los últimos datos aportados por los fabricantes sobre cada fármaco para comprobar la dosis recomendada, la vía y duración de la administración y las contraindicaciones. Es responsabilidad ineludible del médico determinar las dosis y el tratamiento más indicado para cada paciente, en función de su experiencia y del conocimiento de cada caso concreto. Ni los editores ni los directores asumen responsabilidad alguna por los daños que pudieran generarse a personas o propiedades como consecuencia del contenido de esta obra.

**El Editor**

# Índice

Presentación ..... ix

## Parte I. Principios generales de investigación

1. El proceso de la investigación clínica y epidemiológica ..... 3
2. Bases metodológicas de la investigación clínica y epidemiológica ..... 7
3. Ética e investigación ..... 15

## Parte II. Tipos de estudio

4. Clasificación de los tipos de estudio ..... 29
5. Estudios experimentales I: el ensayo clínico aleatorio ..... 33
6. Estudios experimentales II: otros diseños ..... 49
7. Estudios de cohortes ..... 64
8. Estudios de casos y controles ..... 74
9. Estudios descriptivos ..... 87
10. Revisiones sistemáticas ..... 97

## Parte III. Elaboración del protocolo de estudio

11. Protocolo de investigación ..... 109
12. Objetivo del estudio ..... 113
13. Variable de respuesta ..... 121
14. Población de estudio ..... 135
15. Tamaño de la muestra ..... 142

16. Muestreo ..... 155
17. Formación de los grupos de estudio ..... 163
18. Medición de variables ..... 173
19. Selección y definición de las variables ..... 181
20. Diseño de cuestionarios ..... 188
21. Validación de cuestionarios ..... 200
22. Estrategia de análisis ..... 211
23. Sujetos a incluir en el análisis ..... 221
24. Gestión del estudio ..... 232
25. Solicitud de ayuda para la financiación ..... 242

## Parte IV. Interpretación de resultados

26. Interpretación de resultados ..... 249
27. Papel de la estadística ..... 253
28. Potencia de un estudio ..... 265
29. Sesgos ..... 272
30. Confusión y modificación del efecto ..... 280
31. Inferencia causal ..... 292
32. Aplicabilidad práctica de los resultados ..... 298

## Parte V. Publicación

33. Comunicación científica ..... 311
34. El artículo original ..... 316

## Índice

### Anexos

<b>Anexo 1</b>	
Medidas de frecuencia .....	327
<b>Anexo 2</b>	
Medidas de asociación.....	333
<b>Anexo 3</b>	
Sensibilidad y especificidad .....	339
<b>Anexo 4</b>	
Análisis de la concordancia.....	345
<b>Anexo 5</b>	
Medidas del efecto de un tratamiento.....	350
<b>Anexo 6</b>	
Análisis de una revisión sistemática.....	353
<b>Anexo 7</b>	
Guía para la elaboración de un protocolo de estudio.....	358
<b>Anexo 8</b>	
Cálculo del tamaño de la muestra en situaciones especiales .....	361
<b>Anexo 9</b>	
Elección de la prueba estadística.....	364

### Tablas

#### Tabla A

Número de sujetos necesarios para la estimación de una proporción (variable cualitativa).....	371
---	-----

#### Tabla B

Número de sujetos necesarios para la estimación de una media (variable cuantitativa).....	372
---	-----

#### Tabla C

Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde	
---	--

a la proporción menor de las dos que se comparan .....	374
--	-----

#### Tabla D

Número de sujetos necesarios en cada grupo de estudio para estimar un riesgo relativo (RR) .....	385
--	-----

#### Tabla E

Número de sujetos necesarios en cada grupo de estudio para estimar una <i>odds ratio</i> (OR) .....	388
---	-----

#### Tabla F

Número de sujetos necesarios en cada grupo de estudio para la comparación de dos medias mediante la prueba de la t de Student-Fisher .....	390
--	-----

#### Tabla G

Número de sujetos necesarios en cada grupo en estudios de equivalencia.....	393
---	-----

#### Tabla H

Número de sujetos necesarios para la estimación de la concordancia intraobservador.....	395
---	-----

#### Tabla I

Número de sujetos necesarios para la estimación de la concordancia intra e interobservador.....	395
---	-----

#### Tabla J

Número de sujetos necesarios para estimar una asociación entre dos variables cuantitativas mediante el coeficiente de correlación de Pearson .....	396
--	-----

#### Tabla K

Número de sujetos necesarios por grupo para detectar una determinada razón de riesgos en estudios que utilizan análisis de supervivencia .....	397
--	-----

#### Tabla L

Números aleatorios.....	398
Índice alfabético .....	399

**Contenido Online**

Casos estudio

Glosario de términos en español e inglés

Preguntas abiertas de cálculo

Preguntas con respuesta verdadero/falso

Preguntas de múltiple elección

Presentaciones didácticas

Página deliberadamente en blanco

# Presentación

Es un placer para nosotros presentar la cuarta edición de la obra *Métodos de investigación*, cuya primera edición se publicó hace más de 20 años. La idea de escribir un libro surgió a partir de la experiencia de realizar diversos cursos de metodología de la investigación dirigidos a profesionales sanitarios. En aquella primera edición pretendimos escribir un texto sobre fundamentos del método científico que fuera útil para los profesionales que se inician en el campo de la investigación en ciencias de la salud. Con esta filosofía intentamos que el libro fuera lo más sencillo y didáctico posible, centrado en los pasos que se siguen para elaborar un protocolo de investigación, pero a la vez riguroso, como corresponde a un libro de método.

En la segunda edición, se incorporaron nuevos temas, aunque intentando no perder el espíritu de la primera, con la pretensión de que fuera útil, no solamente para los profesionales que se inician en la investigación, sino también para aquellos con experiencia previa en este campo, y para los que quieran aumentar sus capacidades para leer e interpretar críticamente un artículo científico.

La tercera edición incorporaba nuevos temas que iban adquiriendo una importancia mayor en el área de la investigación clínica y epidemiológica, como los aspectos éticos y las revisiones sistemáticas de la literatura científica como herramienta útil para la toma de decisiones. También trataba con mayor profundidad algunos temas de especial relevancia, como por ejemplo los estudios experimentales, que pasaban a ocupar dos capítulos.

Esta cuarta edición consolida las anteriores. Su hilo conductor continua siendo la elaboración de un protocolo de estudio, de forma que aquellos profesionales que se enfrentan al reto de diseñar una investigación puedan seguir paso a paso sus distintas fases, desde los planteamientos iniciales y la definición del objetivo hasta la planificación de la estrategia de análisis. Además de revisar todos los contenidos, especialmente algunos relacionados con la gestión del proyecto, se ha incluido un nuevo capítulo dedicado específicamente a la variable de respuesta dada su importancia y la creciente utilización de variables subrogadas y combinadas en investigación clínica, aspecto que requiere realizar consideraciones especiales. También se han revisado en profundidad y ampliado los aspectos relacionados con los estudios de equivalencia y no-inferioridad, cada vez más frecuentes, que presentan algunos elementos diferenciales relevantes en relación con los estudios habituales de superioridad, especialmente en la determinación del tamaño de la muestra o la interpretación de sus resultados.

Las páginas dedicadas a la interpretación de resultados suponen una parte importante del libro, ya que un investigador debe ser capaz de discutir honestamente sus hallazgos, y los profesionales sanitarios deben poder leer y evaluar críticamente la literatura. Con estas finalidades se presentan los aspectos relacionados con las posibles limitaciones de la validez y la aplicabilidad práctica de los resultados.

Con la finalidad de mantener la utilidad práctica del estudio, los conceptos teóricos se ilustran con más de 300 ejemplos, muchos de ellos revisados y actualizados, que se diferencian en el texto para no entorpecer la lectura de los capítulos.

## Presentación

Como en ediciones anteriores, agradecemos a los profesionales que han participado en nuestros cursos, tanto presenciales como a distancia, sus comentarios y sugerencias, así como al Dr. Amando Martín Zurro, que constantemente nos anima a mantener actualizada esta obra.

Los autores  
Barcelona, 2012

# Parte

1

## Principios generales de investigación

- |  |                                   |
|--|-----------------------------------|
| <b>1</b> El proceso de la investigación clínica<br>y epidemiológica 3          | <b>3</b> Ética e investigación 15 |
| <b>2</b> Bases metodológicas de la investigación<br>clínica y epidemiológica 7 |                                   |

Página deliberadamente en blanco

# Capítulo

1

## El proceso de la investigación clínica y epidemiológica

En la práctica de la medicina, el profesional sanitario debe tomar múltiples decisiones sobre la atención a sus pacientes, para lo cual necesita disponer de información de diferente naturaleza: cuál puede ser la causa de su problema, cuál es su pronóstico, qué estrategia diagnóstica es la más adecuada, qué tratamiento es más eficaz en su situación, qué información y recomendaciones puede proporcionarle para mejorar su estado de salud, etc. Además de tomar en consideración factores como las necesidades y preferencias de los pacientes o las prioridades y recursos que tiene a su alcance, el profesional debe basar sus decisiones en información (evidencia) válida sobre los efectos de las distintas alternativas de actuación. La principal fuente de estas evidencias es la investigación clínica y epidemiológica.

Cada vez existe una mayor presión desde todos los ámbitos para pasar de un proceso de toma de decisiones basado en la tradición, la autoridad o las opiniones de los profesionales considerados expertos, a un modelo basado en evidencias científicas explícitas y contrastables empíricamente, procedentes de la investigación rigurosa, que conduzcan a la obtención de pruebas más objetivas (medicina basada en la evidencia).

Por tanto, la investigación biomédica es necesaria para el progreso de la medicina, ya que proporciona las pruebas en que basar la práctica clínica y mejorar la calidad de la atención que se presta a los pacientes, y permite obtener información sobre la utilidad y eficacia de los procedimientos diagnósticos, terapéuticos y preventivos, así como sobre la etiología, la fisiopatología y los factores de riesgo de las enfermedades y problemas de salud.

La investigación supone también un valor añadido tanto para los profesionales (prestigio profesional y social, estímulo intelectual, sentimiento

de pertenecer a una élite, posibilidad de colaborar con otros colegas y otros profesionales expertos interesados en los mismos temas, etc.) como para los pacientes (beneficios derivados de los resultados de las investigaciones, mayor calidad de la atención, administración de intervenciones más efectivas, mayor supervivencia, mayor calidad de vida, etc.), para la comunidad (mejora de la situación de salud, aumento de la esperanza de vida, etc.) y para el propio sistema sanitario (mayor calidad y eficacia de sus actuaciones, mayor eficiencia en el uso de los recursos, etc.).

### MÉTODO CIENTÍFICO

El objetivo de cualquier ciencia es adquirir conocimientos desde una perspectiva que intenta no abordar solamente acontecimientos y situaciones aisladas, sino la comprensión de fenómenos desde una óptica más generalizada. El problema es elegir un método adecuado que permita conocer la realidad, entendiendo como tal la referida a un determinado problema, evitando catalogar conocimientos erróneos como verdaderos. Es precisamente mediante la aplicación formal de los procedimientos sistemáticos que componen el método científico cómo el investigador pretende comprender, explicar, predecir o controlar determinados fenómenos.

Investigar es algo más que recoger y almacenar información. La investigación nace de la curiosidad y de las inquietudes personales, de la observación de hechos sin explicación lógica aparente o que contradicen las teorías aceptadas. Pero en todos los casos requiere establecer hipótesis y objetivos concretos y utilizar instrumentos de medida precisos y reproducibles con una metodología que permita

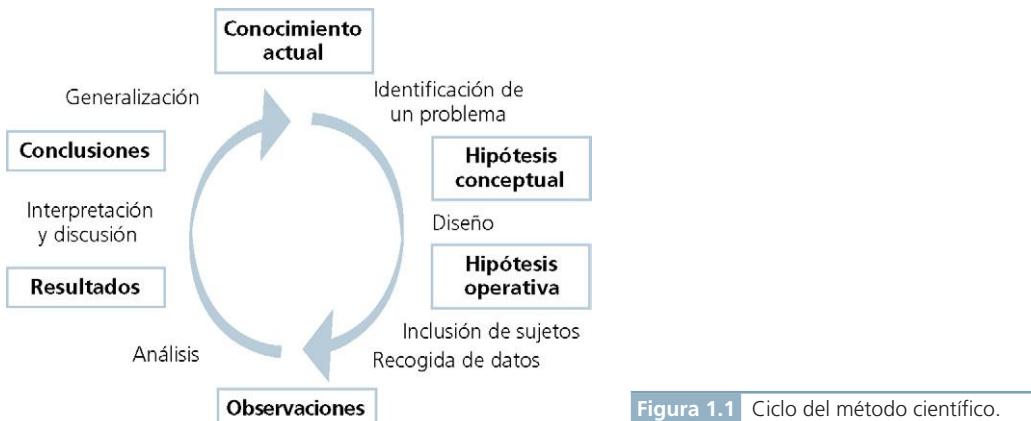


Figura 1.1 Ciclo del método científico.

contrastar empíricamente dichas hipótesis y rechazar o aumentar el grado de corroboración de las teorías aceptadas en ese momento.

Una investigación es un proceso sistemático, organizado y objetivo, destinado a responder a una pregunta. El término *sistemático* significa que se aplica el método científico (fig. 1.1), de manera que, a partir de la identificación de un problema y la revisión de los conocimientos existentes, se formula una hipótesis, se define el objetivo del estudio y se recogen unos datos según un diseño preestablecido, cuyo análisis e interpretación conducen a unas conclusiones, la difusión de las cuales permitirá modificar o añadir nuevos conocimientos a los ya existentes, iniciándose entonces de nuevo el ciclo. Por *organizado* se entiende que todos los miembros de un equipo investigador siguen un mismo protocolo de estudio y aplican las mismas definiciones y criterios a todos los participantes, actuando de forma idéntica ante cualquier duda. La palabra *objetivo* indica que las conclusiones que se obtienen no se basan en impresiones subjetivas, sino en hechos que se han observado, medido y analizado, intentando evitar cualquier prejuicio en la interpretación de los resultados.

El control de las condiciones de investigación es un elemento clave del método científico. Sin embargo, los problemas que interesan a los investigadores son fenómenos complejos y difíciles de medir, que suelen representar los efectos de múltiples factores. Si se pretende aislar las relaciones entre fenómenos, el científico debe intentar controlar los factores que no están siendo investigados de forma directa, lo que resulta más difícil de conseguir en el mundo real que en un laboratorio. Por ello, el método científico aplicado a la investigación en seres humanos no solamente presenta limitaciones morales y éticas, dada la dificultad de controlar los múltiples factores que pueden influir, la complejidad del ser humano

como objeto de investigación y los problemas de medición de algunas de sus funciones.

El resultado de la investigación es conocimiento, pero para que el conocimiento generado por un estudio pase a formar parte de la ciencia, es necesario que se presente a la comunidad científica en una manera que permita juzgar su validez de una forma independiente. Así, las ideas producto de la intuición, la inspiración o la imaginación tienen un alto grado de subjetividad y suelen ser poco fiables. Para cruzar la frontera de la ciencia, deben trasladarse a un proyecto de investigación, que es evaluado por un comité independiente que se encarga de excluir aquello que no tiene sentido o que no puede considerarse ciencia. Si el proyecto pasa este filtro y la investigación se lleva a cabo, debe pasar otro, que es el de su revisión por expertos para decidir si tiene la calidad y el interés suficientes para ser publicada y poderse difundir entre la comunidad científica. Las publicaciones secundarias y la elaboración de revisiones suponen un nuevo filtro, ya que seleccionan los mejores estudios y permiten contrastarlos con otras investigaciones. El paso del tiempo y la experiencia en la aplicación de los resultados se encargan de cribar los que pasan a formar parte del cuerpo de conocimientos vigente (fig. 1.2).

Por tanto, una adecuada gestión del conocimiento debe pasar por la promoción de la investigación útil (producción de información orientada a la resolución de las incertidumbres asociadas a problemas de salud concretos), su adecuada diseminación (transferencia del conocimiento a la práctica profesional) y la formación de los profesionales sanitarios (capacitación técnica para interpretar, comunicar, compartir y utilizar dicho conocimiento). Sin un adecuado fomento de estos aspectos, difícilmente la investigación tendrá el impacto esperado en la práctica de la medicina.



Modificado de Bauer HH (1995). Ethics in science: The knowledge filter  
Disponible en: <http://www.csu.edu.au/learning/eis/hbauer-filter.html>

**Figura 1.2** Filtro del conocimiento.

## PROCESO DE INVESTIGACIÓN

La investigación no es una actividad especialmente difícil, aunque requiere una actitud crítica y la capacidad de pensar con claridad y de una forma organizada, a la vez que ayuda a desarrollarlas. Al contrario de lo que muchos suelen creer, no necesita un extenso conocimiento de técnicas experimentales ni estadísticas, ni el dominio de un amplio vocabulario especializado.

El cuadro 1.1 resume los diez aspectos más importantes que un investigador se plantea y que le servirán de guía en el proceso de elaboración de un proyecto de investigación.

La investigación se inicia a partir de la identificación de un problema o la generación de una buena idea, definiendo la pregunta concreta a la que el estudio pretende responder, expresada habitualmente en forma de hipótesis de trabajo o de objetivo específico. Es importante que el investigador sea capaz de justificar, a partir de una revisión de la situación del conocimiento sobre el tema y del

**Cuadro 1.1 Principales aspectos que se plantean en relación con una investigación**

1. Definir la pregunta con claridad
2. Escoger el diseño idóneo
3. Seleccionar la población de estudio adecuada
4. Calcular el número de individuos necesario
5. Medir las variables con precisión y exactitud
6. Planear la estrategia de análisis
7. Organizar el estudio cuidadosamente
8. Ejecutar el estudio con minuciosidad
9. Interpretar los resultados con precaución
10. Comunicar los resultados con rigor

establecimiento del marco teórico adecuado, la realización del estudio, valorando su pertinencia y su viabilidad. El diseño de la investigación implica, en una primera fase, la selección del tipo de estudio más adecuado para responder a la pregunta planteada y la construcción de su estructura básica. Posteriormente se concretarán los aspectos relacionados con

la población de estudio (criterios de selección de los participantes, número de sujetos necesario, procedimiento de reclutamiento, formación de los grupos de estudio, etc.), las variables (variable de respuesta principal, otras variables que será necesario medir, instrumentos de medida, etc.), la recogida de datos y la estrategia de análisis.

Dado que la utilidad de una investigación depende en gran medida de que sus resultados sean aplicables en la práctica, es fundamental la correcta difusión del trabajo realizado en el ámbito adecuado y su inclusión en las bases de datos internacionales para que pueda ser identificado, localizado y evaluado por los profesionales sanitarios. Por ello, los aspectos relacionados con la comunicación científica son también importantes.

## FINALIDADES DE LA INVESTIGACIÓN

Esquemáticamente, puede considerarse que existen dos grandes categorías de investigación.

La primera corresponde a la investigación que busca ampliar los conocimientos existentes sobre la salud, la enfermedad o el proceso de atención sanitaria (p. ej., estimando la frecuencia con que aparece una enfermedad o describiendo sus características), cuya utilidad principal es la generación de ideas e hipótesis (*investigación descriptiva*). Este tipo de investigación se caracteriza por la ausencia de hipótesis de trabajo previas, y puede ser exclusivamente descriptiva, lo que supone la observación, descripción y catalogación de determinados hechos, o tener una orientación más exploradora, dirigida al descubrimiento de relaciones entre fenómenos. Si se utiliza el método científico para realizar las observaciones, los estudios descriptivos pueden ser muy útiles y suelen ser fundamentales como base del desarrollo de hipótesis.

La segunda categoría corresponde a la investigación dirigida a evaluar las intervenciones realizadas para mejorar la salud, prevenir la enfermedad o impulsar los procesos de atención sanitaria, determinando sus efectos y contrastando las hipótesis de trabajo (*investigación analítica*). Este tipo de investigación parte de una hipótesis previa, y puede tener una orientación explicativa, que busca comprender el porqué de los fenómenos, o bien predictiva, en el sentido de que intenta hacer predicciones fiables y

brindar la posibilidad de controlar determinados problemas, a pesar de que, con los conocimientos y tecnologías disponibles actualmente, escapan a la comprensión absoluta.

Ambas categorías tienen en común el hecho de que el investigador trata de responder a la pregunta de investigación de una forma válida y precisa, diseñando el estudio de manera que disminuyan las probabilidades de existencia de errores que puedan conducirle a una respuesta equivocada. Es mejor tardar en incorporar conocimientos, aunque sean ciertos, que incorporar datos falsos.

## ORIENTACIÓN DE LA INVESTIGACIÓN

Como veremos en los capítulos siguientes, a lo largo de todo el proceso de una investigación deben tomarse múltiples decisiones sobre muchos aspectos, como los criterios de selección de los participantes, la exclusión de determinados grupos de personas, el ámbito de realización o la duración del estudio, el seguimiento de los sujetos, etc., que condicionan la validez del estudio para responder a la pregunta de investigación y su utilidad para extrapolar o generalizar sus resultados.

La mayor parte de los estudios que se realizan en la actualidad, especialmente los dirigidos a evaluar la eficacia de los tratamientos, se diseñan con la finalidad de obtener el mayor grado posible de control de las condiciones de la investigación, es decir, la mayor validez interna posible, por lo que existen importantes limitaciones a la hora de extrapolar sus resultados a la práctica clínica habitual.

Este problema es especialmente relevante si tenemos en cuenta que los resultados de estas investigaciones son la base para la toma de decisiones en la práctica clínica, de manera que, aunque se insiste mucho en la necesidad de basar dichas decisiones en evidencias científicas, en la práctica se dispone de pocos datos de lo que ocurre en realidad en la población a la que se desea aplicar los resultados y en las condiciones reales de la práctica diaria.

Además, los principales retos actuales de la investigación clínica, como las patologías crónicas, las actividades preventivas o los efectos a largo plazo de las intervenciones en poblaciones con problemas complejos o con comorbilidad, requieren una orientación más pragmática del diseño de los estudios.

# Capítulo

# 2

## Bases metodológicas de la investigación clínica y epidemiológica

Esquemáticamente, la pregunta o hipótesis que se formula en un estudio puede hacer referencia a la frecuencia con que aparece una enfermedad o a la descripción de sus características (*estudios descriptivos*), o bien a la estimación del efecto de un factor de estudio (una exposición o una intervención) sobre una enfermedad o una variable de respuesta (*estudios analíticos*). En todos los casos, el investigador trata de responder a la pregunta de forma válida y precisa. En otras palabras, el estudio se diseña para evitar cualquier error en la respuesta al objetivo planteado.

### TIPOS DE ERROR

En cualquier estudio se han de considerar dos tipos de posibles errores: el error aleatorio y el error sistemático.

#### Error aleatorio

El error aleatorio se debe al azar. Ocurre tanto por el hecho de que se trabaja con muestras de individuos, y no con poblaciones enteras, como por la variabilidad inherente al proceso de medición de variables, ya sea por el instrumento de medida que se utiliza, por la propia variabilidad biológica o por la debida al observador.

Supongamos que se desea conocer el porcentaje de historias clínicas en las que están registrados los antecedentes familiares de cardiopatía isquémica. Para ello, se selecciona una muestra de historias clínicas, determinando el tanto por ciento que contiene la información de interés e infiriendo que el porcentaje observado en la muestra es el mismo que se encontraría si se hubieran analizado todas las historias. Sin embargo, existe la posibilidad de equivocarse,

simplemente porque, por azar, se ha escogido una muestra de historias que no refleja de forma exacta el verdadero tanto por ciento. Esta posibilidad es tanto mayor cuanto menor es el tamaño de la muestra estudiada. Si sólo se hubieran seleccionado diez historias clínicas, y en cuatro de ellas estuvieran registrados los antecedentes familiares de cardiopatía isquémica, podría inferirse que el porcentaje de historias en que constan dichos antecedentes es del 40%. Si se repitiera el estudio seleccionando otra muestra de diez historias, es muy posible que el tanto por ciento observado fuera diferente del 40% (una sola historia más o menos haría variar dicho porcentaje en un 10%). Es decir, existe una gran variabilidad inherente al muestreo. Esta variabilidad sería menor si, en lugar de diez historias, se hubieran seleccionado cien (una historia haría variar el resultado en un 1%), e incluso podría desaparecer si se estudiaran todas las historias clínicas. Éste es un ejemplo de error aleatorio debido al hecho de que se trabaja con muestras. La mejor estrategia para reducirlo sería aumentar el tamaño de la muestra.

El error aleatorio también puede producirse al medir las variables. Por ejemplo, las cifras de presión arterial presentan una gran variabilidad. Una forma de reducirla es medir la presión arterial en más de una ocasión y utilizar algún promedio de las diferentes medidas. Además, existen otras fuentes de variación (momento del día, condiciones de la medición, tipo de esfigmomanómetro...) que también deben ser controladas mediante la estandarización de las condiciones de la medición y el entrenamiento de los observadores.

El error aleatorio está muy relacionado con el concepto de *precisión*. Una estimación o una medida es tanto más precisa cuanto menor es el componente de error aleatorio.

## Error sistemático

Un error sistemático, o sesgo, es un error en el diseño del estudio, ya sea en la selección de los sujetos (*sesgo de selección*) o en la medición de las variables (*sesgo de información*), que conduce a una estimación incorrecta o no válida del efecto o parámetro que se estudia.

### Sesgo de selección

El sesgo de selección puede aparecer al elegir una muestra que no represente de forma adecuada a la población de estudio o al formar los grupos que se van a comparar.

**Ejemplo 2.1.** Supongamos que se desea estimar la prevalencia de una enfermedad en los sujetos adultos residentes en un municipio (población de estudio). Para ello, se estudian los cien primeros sujetos que acuden a la consulta en un centro de salud (muestra estudiada). Obviamente, los sujetos que acuden a la consulta no son representativos de la población del municipio, ya que tienen una mayor probabilidad de estar enfermos, por lo que se obtendría una estimación sesgada de la prevalencia de la enfermedad.

**Ejemplo 2.2.** Supongamos que se estudia la relación entre la aparición de cáncer de pulmón y la exposición previa al humo del tabaco. Se observa que el 90% de los sujetos con cáncer de pulmón ingresados en un hospital (casos) eran fumadores. Supongamos, además, que los individuos del grupo control se eligen entre los pacientes ingresados en el servicio de neumología, de forma que la mayoría de ellos padece limitación crónica al flujo aéreo (LCFA), y que el 88% había estado expuesto al tabaco. Ante estos resultados se podría concluir erróneamente que no existe asociación entre el tabaco y el cáncer de pulmón, ya que el tanto por ciento de fumadores es muy similar en ambos grupos. Sin embargo, la selección del grupo control no ha sido muy afortunada, ya que la LCFA está relacionada

con el tabaco, lo que enmascara su asociación con el cáncer de pulmón. En general, si se eligen como controles personas con enfermedades relacionadas positiva o negativamente con la exposición, se está introduciendo un sesgo de selección.

Los sesgos de selección también se pueden producir durante el seguimiento de los participantes si la probabilidad de desarrollar la enfermedad entre los sujetos que se pierden es diferente en cada uno de los grupos.

**Ejemplo 2.3.** Consideremos un estudio que compara dos intervenciones (A y B), administradas cada una de ellas a un grupo de 100 sujetos, y que se producen 20 pérdidas durante el seguimiento en cada uno de los grupos (tabla 2.1). El porcentaje de éxitos con ambas intervenciones es del 50% entre los que finalizan el estudio, por lo que se podría concluir que poseen la misma eficacia. Sin embargo, entre las pérdidas, sólo el 1% de los sujetos que recibieron la intervención A fue clasificado como éxito, mientras que con la intervención B lo fue un 50%. Cuando se calculó el tanto por ciento total de éxitos resultó que la intervención B fue la más eficaz. Así pues, aunque el porcentaje de pérdidas sea el mismo en ambos grupos, se puede producir un sesgo de selección, ya que sus características son diferentes.

Otra situación que puede producir un sesgo de selección es la existencia de preguntas no contestadas, lo que suele ocurrir en las encuestas transversales.

**Ejemplo 2.4.** Supongamos un estudio en el que se desea determinar la prevalencia de consumo de tabaco entre los profesionales sanitarios de una determinada zona geográfica. Para ello, se selecciona una muestra aleatoria de 500 profesionales, a los que se les envía un cuestionario. Contestan 300 profesionales, de los que 75 (25%) reconocen que fuman. Si la prevalencia de consumo de tabaco entre quienes no han respondido es diferente de la que existe entre los que sí lo han hecho, la cifra

**Tabla 2.1 Sesgo de selección debido a pérdidas de seguimiento (ejemplo 2.3)**

	Intervención A	Intervención B
Eficacia entre los sujetos que finalizan el estudio	40/80 (50%)	40/80 (50%)
Eficacia entre los sujetos que no lo terminan	1/20 (5%)	10/20 (50%)
Total	41/100 (41%)	50/100 (50%)

del 25% es una estimación sesgada de la verdadera prevalencia. En caso contrario, podría asumirse que las preguntas no contestadas no están relacionadas con el fenómeno de estudio (podrían considerarse aleatorias), de forma que la cifra observada sería una estimación no sesgada, aunque se habría producido una pérdida de precisión en la estimación debido al menor número de respuestas.

La única manera de asegurar que las pérdidas durante el seguimiento o las preguntas no contestadas no introducen un error sistemático en los resultados es evitar que se produzcan, o bien obtener información suplementaria que permita evaluar si los sujetos que se pierden o que no contestan difieren de los que finalizan el estudio.

## Sesgo de información

El sesgo de información se produce cuando las mediciones de las variables de estudio son de mala calidad o son sistemáticamente desiguales entre los grupos. Las principales fuentes de este error son la aplicación de pruebas poco sensibles y/o específicas para la medición de las variables, o de criterios diagnósticos incorrectos o distintos en cada grupo, así como las imprecisiones y omisiones en la recogida de los datos.

**Ejemplo 2.5.** Si en un estudio se mide el peso de los sujetos con una báscula mal calibrada, evidentemente, se obtendrán medidas incorrectas, ya que se utiliza un instrumento de medida inadecuado.

**Ejemplo 2.6.** Consideremos un estudio cuyo objetivo es comparar el peso en función del sexo con una báscula bien calibrada. Sin embargo, la medición en los hombres se hace sin zapatos y en ropa interior, mientras que el responsable de pesar a las mujeres sigue un criterio diferente y efectúa la medición con zapatos y en ropa de calle. El problema de este estudio es que se aplica un mismo instrumento de medida de forma diferente, según el grupo de estudio.

**Ejemplo 2.7.** Si se quiere estudiar si hay una asociación entre la ingestión de alcohol y la hipertensión, y en los hipertensos se obtiene la información a partir de una entrevista personal, mientras que en los individuos del grupo control se obtiene de las historias clínicas, es de esperar que en el primer grupo la información sea más exacta y sistemáticamente diferente de la obtenida en el grupo control, lo que sesgará los resultados. El problema de este estudio es que se usan instrumentos diferentes en cada grupo.

Por ello, las variables deben medirse con el instrumento adecuado, validado y bien calibrado, que debe aplicarse de la misma forma en todos los participantes del estudio.

Los errores sistemáticos, ya sean de selección o de información, a diferencia de lo que ocurre con el error aleatorio, no se atenúan al aumentar el tamaño de la muestra. De hecho, aunque se incluyan más individuos, lo único que se logra con ello es perpetuar el sesgo. Además, un error de estas características, una vez introducido, es casi imposible de enmendar en la fase de análisis.

El error sistemático va muy ligado al concepto de *validez*. La estimación de un parámetro o de un efecto se considera válida si representa el verdadero valor del fenómeno que se desea medir.

## GRUPO CONTROL

En los estudios analíticos se pretende estimar la asociación o el efecto de un factor de estudio sobre una variable de respuesta; por ejemplo, qué reducción del colesterol sérico (variable de respuesta) se consigue con un determinado tratamiento hipolipemiantes (factor de estudio). En estos diseños, además del grupo que está expuesto al factor de estudio, es necesario utilizar un grupo control que sirva de referencia sobre lo que ocurre en los sujetos no expuestos a dicho factor, de forma que puedan compararse los resultados obtenidos en ambos grupos. El siguiente ejemplo demuestra la necesidad del grupo control en estos estudios.

**Ejemplo 2.8.** En la década de 1950, un tratamiento para la angina de pecho que gozaba de gran popularidad entre los cirujanos era la ligadura de la arteria mamaria interna. Se pensaba que con ello se aumentaba el flujo sanguíneo hacia el miocardio. Cobb et al (1959) publicaron un estudio controlado, doble ciego, en el que a los individuos del grupo experimental se les practicaba la ligadura, mientras que a los del grupo control se les practicaba una incisión superficial sin ligar la arteria, con el fin de que ni los pacientes ni los investigadores que evaluaban los resultados pudieran conocer si pertenecían a un grupo o al otro. En este estudio se observó la recuperación del 32% de los pacientes a quienes se había ligado la arteria y del 43% de los controles. Si no hubiera existido un grupo control se hubiera podido llegar a la errónea conclusión de que la ligadura era eficaz.

La función de un grupo control es estimar cuál sería el valor de la variable de respuesta en el caso de

que no existiera asociación con el factor de estudio, por lo que debe formarse de manera que sea comparable con el grupo de estudio en todos los aspectos excepto en la exposición o el tratamiento que recibe. Así, en un ensayo clínico serviría para determinar la respuesta esperada en ausencia de la intervención, o en un estudio de casos y controles, para estimar el tanto por ciento esperado de sujetos expuestos si no existiera asociación con la enfermedad.

### Efectos que controlar

Si no se utiliza un grupo control, es difícil saber si los resultados obtenidos son debidos al factor de estudio o a otros efectos que no se han controlado. A continuación se describen los principales efectos que se han de controlar cuando se diseña un estudio.

### Efecto Hawthorne

El efecto Hawthorne es una respuesta inducida por el conocimiento de los participantes de que están siendo estudiados. La participación en un estudio puede cambiar el comportamiento de los individuos tanto del grupo de estudio como del de control. El hecho de explicarles en qué consiste la experiencia, unido a que durante un tiempo deban visitar al médico con más frecuencia, puede hacer que los participantes alteren sus hábitos y obtengan una respuesta que no puede ser atribuida al factor de estudio.

Se describió por primera vez durante la década de 1920, cuando la Western Electric Company llevó a cabo una serie de experiencias en su fábrica Hawthorne de Chicago con el fin de determinar el efecto de la iluminación en la producción. Los grupos control trabajaron bajo una iluminación constante, mientras que en los grupos experimentales era variable, aumentada o disminuida. El resultado fue que la producción se incrementó no sólo en los grupos de estudio, con independencia de la intensidad de la iluminación, sino también en los grupos control. Parecía claro, pues, que el simple hecho de que los trabajadores sabían que eran vigilados supuso un aumento de la producción.

En los estudios sin grupo de comparación es muy difícil discernir entre el efecto de la intervención y el debido al hecho de saberse estudiado. Con un grupo de comparación, este efecto no se elimina, pero es de esperar que ocurra por igual en ambos grupos y, de este modo, es posible conocer cuál es el verdadero efecto de la intervención.

**Ejemplo 2.9.** Un ensayo clínico evaluaba dos estrategias para reducir el número de peticiones de laboratorio y radiología solicitadas por los médicos residentes de primer año (Martin et al, 1980). Un

tercer grupo, al que no se le practicaba ninguna intervención, sirvió de control. Los residentes tenían conocimiento de la existencia del estudio y de cuál era su objetivo. Cuando se comparó el número de peticiones al inicio y al final del estudio de cada grupo, se obtuvieron diferencias estadísticamente significativas en todos ellos, incluido el de control, lo que sugería la existencia de un efecto Hawthorne. Sin embargo, al comparar los grupos entre ellos, se observó que los dos grupos de estudio tuvieron un descenso significativo en el número de peticiones en relación con el grupo control, lo que sugiere que las estrategias puestas en marcha fueron efectivas.

### Efecto placebo

El efecto placebo se puede definir como la respuesta que se produce en una persona enferma como consecuencia de la administración de un tratamiento, pero que no puede considerarse como un efecto específico del mismo. Por esta razón, cuando una enfermedad no tiene un tratamiento activo conocido o ampliamente aceptado y se quiere evaluar la eficacia de un nuevo fármaco, es importante que al grupo control se le administre un tratamiento lo más parecido posible al nuevo fármaco en todas las características, excepto en la actividad farmacológica, con el fin de controlar el efecto placebo. El hecho de tener un grupo control al que se administre un placebo permite aislar el verdadero efecto debido a la intervención. Sin grupo de comparación es imposible diferenciar qué parte se debe al efecto del fármaco y cuál a otros.

### Regresión a la media

Se entiende por regresión a la media la tendencia de los individuos que tienen un valor extremo de una variable a presentar valores más cercanos a la media de la distribución cuando esta variable se mide por segunda vez.

Los tratamientos, o intervenciones en general, se ensayan cuando una variable que expresa una enfermedad o un factor de riesgo presenta valores fuera de lo común, por lo que es posible que en subsiguientes mediciones los valores sean menos extremos, incluso en ausencia de una intervención eficaz. Así puede suceder, por ejemplo, cuando se seleccionan sujetos porque en una visita presentan un valor elevado de la presión arterial, y sin embargo, en visitas posteriores algunos tienen cifras más bajas simplemente por la variabilidad de la presión arterial, y no por un efecto del tratamiento en estudio.

**Ejemplo 2.10.** En la primera visita del Hypertension Detection and Follow-up Program (1979) para determinar si los posibles candidatos cumplían los criterios de selección, un grupo de participantes tenía una media de tensión diastólica de 122,1 mmHg. En la siguiente visita, la media de tensión diastólica de estos mismos sujetos había disminuido a 114,7 mmHg, aunque la intervención aún no había empezado.

La regresión a la media es una fuente de muchos errores en la interpretación de los resultados de un estudio. Cuando los individuos se seleccionan según una característica que varía con el tiempo, el efecto que se halle en las siguientes mediciones puede deberse a la variabilidad de la medida y no al efecto de la intervención. Cuando se dispone de un grupo de comparación, este fenómeno no desaparece, pero se controla, ya que es de suponer que sucederá en ambos grupos por igual.

## Evolución natural

Cuando el curso habitual de una enfermedad tiende hacia su resolución, los esfuerzos terapéuticos pueden coincidir con la recuperación observada, pero no ser su causa. De ahí la importancia de comparar siempre los resultados obtenidos en un estudio con los del grupo control.

La figura 2.1 resume la función del grupo control en distintas situaciones. La columna A muestra el efecto total observado en el grupo de estudio y sus respectivos componentes. La columna B corresponde a la situación en la que el grupo control recibe un placebo, por lo que es posible aislar perfectamente el efecto de la intervención. La columna C representaría la situación en la que, por las propias características

de la intervención (p. ej., la educación sanitaria), es imposible obtener un grupo control con placebo. Los integrantes del grupo control saben que están participando en un estudio y son controlados con la misma pauta de visitas y exploraciones, pero no se les administra la intervención. Por último, la columna D ilustra la situación de un estudio en el que el grupo control no sabe que está siendo estudiado y, por tanto, sólo se controla la evolución natural de la enfermedad.

## FACTORES DE CONFUSIÓN

Los estudios analíticos implican una comparación entre grupos. El principal requisito para que esta comparación sea válida es que los grupos sean similares en relación con las características que influyen sobre los resultados. A menudo, estas variables se distribuyen de modo desigual y es necesario corregir estas diferencias. Tomemos un ejemplo sencillo: los resultados de una investigación muestran que la incidencia de la LCFA es mucho mayor en hombres que en mujeres: ¿significa esta diferencia que el sexo es un factor de riesgo, o bien se puede explicar por el diferente hábito tabáquico entre hombres y mujeres? En este último caso, el tabaco sería un factor de confusión de la relación entre el sexo y la incidencia de la LCFA.

El fenómeno de confusión aparece cuando la asociación observada entre el factor de estudio y la variable de respuesta puede ser total o parcialmente explicada por una tercera variable (factor de confusión) o, por el contrario, cuando una asociación real queda enmascarada por este factor. Para que una variable se considere de confusión debe estar asociada tanto al factor de estudio como a la respuesta, y no debe ser un paso intermedio en la relación entre ellos.

**Ejemplo 2.11.** Supongamos un estudio hipotético de la asociación entre el consumo de café y el cáncer de laringe (tabla 2.2). A partir de los datos totales se estima que el porcentaje de expuestos entre los enfermos es del 58%, superior al 42% observado entre los controles, lo que sugiere una asociación entre el consumo de café y el cáncer de laringe. Sin embargo, al analizar los datos en subgrupos o estratos en función de si los sujetos eran fumadores o no, se observa que el tanto por ciento de sujetos expuestos al consumo de café es el mismo en los casos y en los controles, y desaparece la asociación. Así pues, la variable «fumador» actúa como un factor de confusión. Es un factor de riesgo del cáncer de laringe (respuesta) y está relacionado con la variable en estudio (consumo de café), ya que la proporción

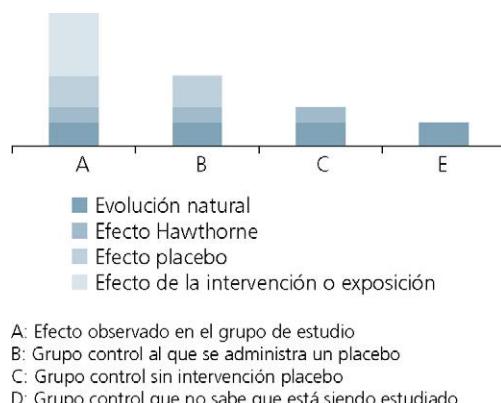


Figura 2.1 Función del grupo control.

**Tabla 2.2 Control de un factor de confusión. Estudio hipotético de la asociación entre consumo de café y cáncer de laringe (ejemplo 2.11)**

	Café (+)	Café (-)	Total
<b>Datos globales</b>			
Enfermos con cáncer de laringe	175 (58%)	125 (42%)	300 (100%)
Controles	125 (42%)	175 (58%)	300 (100%)
<b>Fumadores</b>			
Enfermos con cáncer de laringe	150 (75%)	50 (25%)	200 (100%)
Controles	75 (75%)	25 (25%)	100 (100%)
<b>No fumadores</b>			
Enfermos con cáncer de laringe	25 (25%)	75 (75%)	100 (100%)
Controles	50 (25%)	150 (75%)	200 (100%)

de controles fumadores que toman café habitualmente (75:100) es mucho mayor que la proporción de los no fumadores (50:200).

Un factor de confusión puede sobreestimar o infraestimar un efecto.

**Ejemplo 2.12.** Consideremos un estudio en el que se evalúa la asociación entre el ejercicio físico y el infarto agudo de miocardio (IAM), partiendo de la hipótesis de que a más ejercicio físico, menor es el riesgo de sufrir un IAM. Una variable que podría ejercer de factor de confusión sería la edad. Por un lado, es más probable que los más jóvenes realicen más ejercicio (relación con la variable de estudio) y, por otro, que tengan un riesgo menor de desarrollar un IAM (relación con la enfermedad). Si se cumplen estas condiciones, la edad actuaría como un factor de confusión que tendería a sobreestimar el efecto protector del ejercicio físico sobre el IAM.

**Ejemplo 2.13.** Siguiendo con el ejemplo anterior, el sexo también podría ser un factor de confusión, ya que los hombres suelen realizar más ejercicio físico que las mujeres y, además, presentan un riesgo mayor de padecer un IAM. Así pues, la asociación protectora del ejercicio físico sobre el IAM quedaría infraestimada si no se tuviera en cuenta la diferencia en la proporción de hombres y mujeres que pudiera haber en la muestra.

En la situación más extrema, un factor de confusión puede invertir la dirección de una asociación: es la llamada paradoja de Simpson.

**Ejemplo 2.14.** Rothman (1986) presenta el siguiente ejemplo: «supongamos que un hombre entra en una tienda para comprarse un sombrero y encuentra una estantería con 30, 10 de ellos negros y 20 grises. Descubre que 9 de los 10 sombreros negros le van bien, pero que de los 20 grises sólo le van bien 17. Por tanto, toma nota de que la proporción de sombreros negros que le van bien es del 90%, mientras que la de los grises es sólo del 85%. En otra estantería de la misma tienda encuentra otros 30 sombreros, 20 negros y 10 grises. En ella, 3 (15%) de los sombreros negros le van bien y sólo 1 (10%) de los grises. Antes de que escoja un sombrero, la tienda cierra y él decide volver al día siguiente. Durante la noche, un empleado ha puesto todos los sombreros en una única estantería: ahora hay en ella 60 sombreros, 30 de cada color. El cliente recuerda que el día anterior la proporción de sombreros negros que le iba bien era superior en ambas estanterías. Hoy se da cuenta de que, aunque tiene delante los mismos sombreros, una vez mezclados, sólo el 40% de los sombreros negros (12 de 30) le va bien, mientras que, de los grises, le va bien el 60% (18 de 30)». Aunque este curioso cambio es conocido como la paradoja de Simpson, no es realmente ninguna paradoja. Este fenómeno es análogo al de confusión, que puede distorsionar una asociación, incluso hasta el punto de cambiar su dirección.

La identificación de las variables que pueden ser potenciales factores de confusión es, a menudo, difícil. El conocimiento teórico sobre los mecanismos causales de la respuesta y la revisión de estudios que hayan abordado un objetivo similar serán de gran ayuda para conocer qué variables pueden ser

potenciales factores de confusión. Esta identificación es importante, ya que permitirá controlar el efecto de confusión en la fase de análisis. En el ejemplo 2.11, la determinación del consumo de tabaco (potencial factor de confusión) permitió realizar un análisis en estratos (fumadores y no fumadores), de tal forma que se controlaba su efecto. Ésta es una diferencia con los sesgos, que, una vez introducidos, no pueden controlarse en el análisis.

## VALIDEZ INTERNA Y VALIDEZ EXTERNA

La *valididad interna* se refiere al grado en que los resultados de un estudio son válidos (libres de error) para la población que ha sido estudiada. Los errores sistemáticos y los factores de confusión afectan a la validez interna de un estudio.

Por el contrario, la *valididad externa* se refiere al grado en que los resultados de un estudio pueden ser extrapolados a poblaciones distintas a la estudiada o generalizados a otras más amplias.

La primera condición para que unos resultados puedan ser generalizados es que tengan validez interna. Algunos estudios se diseñan pensando en conseguir la máxima validez interna, aunque sea a costa de grandes limitaciones de la validez externa (estudios explicativos). En otras ocasiones, se desea tener la máxima validez externa (estudios pragmáticos), pero esta actitud no debe llevar a sacrificar la validez interna del estudio.

La generalización o extrapolación de los resultados no se hace a partir de un único estudio, sino que debe tener en cuenta si hipótesis similares, contrastadas en poblaciones distintas y en ámbitos y áreas geográficas diferentes, conducen a resultados similares. Además, debe apoyarse en criterios de plausibilidad biológica que traten de explicar el mecanismo por el cual se produce un efecto determinado. Así pues, la generalización depende de la consistencia de resultados entre diferentes estudios y de mecanismos plausibles que expliquen de forma razonable un efecto, lo que a menudo implica la emisión de juicios de valor por parte de los propios investigadores.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

Cobb LA, Thomas GI, Dillard DH, Merendino KA, Bruce RA. An evaluation of internal-mammary-artery ligation by a double-blind technic. *N Engl J Med.* 1959;260:1115-8.

Hypertension Detection and Follow-up Program Cooperative Group. The Hypertension Detection and Follow-up Program: a progress report. *Cir Res.* 1979;40(Supl 1):106-9.  
Martin AR, Wolf MA, Thibodeau LA, Dzau V, Braunwald E. A

trial of two strategies to modify the test ordering behavior of medical residents. *N Engl J Med.* 1980;303:1330.  
Rothman KJ. *Modern epidemiology*. Boston: Little, Brown and Company; 1986.

## BIBLIOGRAFÍA

Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol.* 2005;34:215-20.  
Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health.* 2004;58:635-41.  
García Alonso F, Guallar E, Bakke OM, Carné X. El placebo en ensayos clínicos con medicamentos. *Med Clin (Barc).* 1997;109:797-801.

Gerhard T. Bias: considerations for research practice. *Am J Health Syst Pharm.* 2008;65:2159-68.  
Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol.* 2003;3:28.  
Greenland S, Morgenstern H. Confounding in health

research. *Annu Rev Public Health.* 2001;22:189-212.  
Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet.* 2002;359:57-61.  
Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet.* 2002;359:248-52.  
Guallar E, Jiménez FJ, García Alonso F, Bakke OM. La regresión a la media en la investigación y práctica clínica.

- Med Clin (Barc). 1997;109: 23-6.
- Hartung DM, Touchette D. Overview of clinical research design. Am J Health Syst Pharm. 2009;66:398-408.
- Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. N Engl J Med. 2001;344:1594-602.
- Hrobjartsson A. What are the main methodological problems in the estimation of placebo effects? J Clin Epidemiol. 2002;55:430-5.
- Moerman DE, Jonas WB. Deconstructing the placebo effect and finding the meaning response. Ann Intern Med. 2002;136:471-6.
- Morton V, Torgerson DJ. Effect of regression to the mean on decision making in health care. BMJ. 2003;326(7398):1083-4.
- Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia: Lippincott, Williams and Wilkins; 2008.
- Yudkin PL, Stratton IM. How to deal with regression to the mean in intervention studies. Lancet. 1996;347:241-2.

# Capítulo

# 3

## Ética e investigación

La investigación biomédica es necesaria para el progreso de la medicina porque proporciona las pruebas sobre las que basar la práctica clínica, contribuyendo así a mejorar la calidad de la atención que se presta a los pacientes, obteniendo información sobre la utilidad y eficacia de los procedimientos diagnósticos, terapéuticos y preventivos, así como sobre la etiología, fisiopatología y factores de riesgo de las enfermedades y problemas de salud. Dados los importantes beneficios que todo ello supone para la sociedad, resulta obvio que la investigación debe considerarse una obligación moral de las profesiones sanitarias. Por lo tanto, aunque suelen discutirse frecuentemente los aspectos éticos relacionados con la realización de investigaciones sanitarias, también habría que preguntarse si es ético no investigar para intentar resolver las lagunas de conocimiento, dudas e incertidumbres que se plantean en el quehacer diario del profesional.

La investigación implica en muchos casos la participación de seres humanos, con los riesgos, molestias e incomodidades que suele suponer para ellos, por lo que es frecuente que se planteen cuestiones éticas y legales, a veces difíciles de resolver. El campo de la **ética de la investigación** está dedicado al análisis sistemático de estas cuestiones para asegurar que los participantes en un estudio estén protegidos y, en último término, que la investigación clínica se conduce de manera que sirva a las necesidades de esos participantes, así como a las de la sociedad en su conjunto (Weijer et al, 1997).

Por otro lado, el resultado de la investigación es conocimiento, el cual debe difundirse de forma adecuada entre la comunidad científica, lo que habitualmente significa su publicación en una revista y su inclusión en las bases de datos bibliográficas. En todas las fases de la comunicación científica son

múltiples los conflictos de intereses que pueden presentarse. Dado que la literatura científica es la principal fuente de evidencias que se utiliza en la toma de decisiones y en la elaboración de las guías de práctica clínica y las recomendaciones para la atención de los pacientes, el impacto que estos conflictos pueden tener sobre la práctica es muy importante.

### ÉTICA E INVESTIGACIÓN CLÍNICA

La ética de la investigación tiene dos componentes esenciales: la selección y el logro de *fines* moralmente aceptables, y la utilización de *medios* moralmente aceptables para alcanzar esos fines.

La primera premisa que se debe tener en cuenta es que lo que no es científicamente correcto es éticamente inaceptable. Pero el rigor científico, aunque es condición necesaria, no es suficiente, ya que no todo lo científicamente correcto es aceptable desde el punto de vista ético. La historia muestra numerosos ejemplos de estudios éticamente inaceptables. Aunque suelen citarse los experimentos realizados por médicos nazis durante la Segunda Guerra Mundial, existen otros muchos ejemplos.

**Ejemplo 3.1.** El estudio de Tuskegee (EE.UU.), iniciado en 1932, consistía en el seguimiento de una cohorte de más de 400 sujetos de raza negra con sífilis no tratada, que se comparaba con un grupo de 204 sujetos sin sífilis, para estudiar la historia natural de la enfermedad. Aunque no existía un tratamiento eficaz al inicio del estudio, el seguimiento continuó sin que los pacientes recibieran tratamiento incluso después de que se hubiera demostrado la eficacia

de la penicilina –hecho del que no se informó a los participantes– y a pesar de que los datos mostraban claramente un peor pronóstico y un aumento de la mortalidad entre los sujetos infectados. El estudio se finalizó en 1972 debido a la presión social que se creó cuando el público general tuvo conocimiento del mismo. Ante las críticas, los investigadores argumentaron que, dado que los participantes eran afroamericanos pobres, aunque no hubieran participado en el estudio tampoco habrían tenido acceso al tratamiento.

**Ejemplo 3.2.** Willowbrook era una institución del estado de Nueva York para personas con deficiencias mentales donde se realizaron diferentes estudios para analizar la historia natural de la hepatitis y los efectos de la gammaglobulina sobre ella. Los sujetos de estudio eran niños a los que se infectaba deliberadamente con el virus. Los investigadores defendían esta actuación argumentando que la mayoría de los niños internados en el centro acababa adquiriendo la infección de forma espontánea.

**Ejemplo 3.3.** Ejemplos más recientes son los estudios, publicados a finales de la década de 1990, realizados en países en vías de desarrollo, sobre la prevención de la transmisión vertical del virus de la inmunodeficiencia humana (VIH). Prácticamente todos los estudios utilizan un grupo control al que administran placebo a pesar de que existen pruebas sobre la eficacia de determinados tratamientos y de la existencia de recomendaciones sobre su uso en países occidentales (Angell, 1997; Lurie y Wolfe, 1997). De forma similar, en un ensayo clínico aleatorio realizado en Uganda sobre la profilaxis de la tuberculosis en adultos VIH positivos con prueba positiva de la tuberculina (Whalen et al, 1997), el grupo control recibió placebo a pesar de la existencia de recomendaciones elaboradas por los Centers for Disease Control (CDC) desde 1990.

Como respuesta a estos y otros casos se han elaborado diferentes códigos éticos y normativas legales para guiar la realización de investigaciones con seres humanos. Los dos pilares fundamentales son el **Código de Núremberg**, elaborado en 1947, tras la Segunda Guerra Mundial (<http://www.cirp.org/library/ethics/nuremberg/>; versión traducida al castellano en <http://www.pcb.ub.edu/bioeticaidret/archivos/norm/CodigoNuremberg.pdf>) y la **Declaración de Helsinki**, elaborada por la Asociación Médica Mundial en 1964 y revisada en diferentes ocasiones, que se considera la principal referencia mundial de investigación biomédica (<http://www.wma.net/en/30publications/10policies/b3/index.html>); versión traducida al castellano en: <http://www.wma.net/es/30publications/10policies/b3/>).

La Declaración de Helsinki presenta los principios éticos básicos de toda investigación médica en seres humanos, incluida la investigación de material humano y de información identificables, y los aplicables cuando ésta se combina con la atención médica. Esta declaración se ha revisado en diversas ocasiones: Tokio (1975), Venecia (1983), Hong Kong (1989), Somerset West (1996), Edimburgo (2000) y Seúl (2008). En la última revisión se han incorporado aspectos como la inclusión, entre los deberes del médico en una investigación, además de la protección de la vida, la salud, la dignidad, la integridad y la intimidad de los participantes, la confidencialidad de la información personal o el registro de todo ensayo clínico en una base de datos de acceso público antes del reclutamiento del primer sujeto.

Dado que estos códigos eran difíciles de interpretar y aplicar en ocasiones, y se iban produciendo nuevos casos de investigaciones no éticas, como el de Willowbrook (ejemplo 3.2) y especialmente el de Tuskegee (ejemplo 3.1), el congreso de EE.UU. creó en 1974 la Comisión Nacional para la Protección de Sujetos Humanos en la Investigación Biomédica y Conductual, para que elaborara un documento que enunciara los principios éticos básicos para formular, criticar e interpretar reglas o normas específicas de aplicación práctica en la investigación clínica en seres humanos. En 1978 esta comisión presentó el documento sobre principios éticos y pautas para la protección de sujetos humanos de la investigación conocido como **Informe Belmont** (<http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html>; versión traducida al castellano en <http://www.pcb.ub.edu/bioeticaidret/archivos/norm/InformeBelmont.pdf>), origen de la moderna teoría ética de la investigación clínica, en el que se definen los tres «principios éticos básicos»: el respeto por las personas, la beneficencia y la justicia.

Una de las principales aportaciones del Informe Belmont, como comenta De Abajo (2001), fue el reconocimiento de que la investigación clínica es una actividad primariamente cognoscitiva, y no beneficiante como la práctica clínica, y por lo tanto se requería una justificación diferente de la que se había dado hasta ese momento. La legitimación ética y social de la investigación biomédica en seres humanos debería venir dada por el hecho de que la práctica clínica, para poder considerarse correcta, debe estar basada en pruebas objetivas y no, o al menos no tanto o primariamente, en opiniones y experiencias personales. La existencia de pruebas científicas distingue las prácticas validadas de las

no validadas o simplemente empíricas, y el procedimiento de validación es precisamente la investigación clínica, por lo que ésta se justificaba en la medida en que era una condición de posibilidad de una práctica clínica correcta y, por tanto, ética. Posteriormente, la investigación debe tener su propia validación ética, cumpliendo con los principios y normas adecuados, pero antes de la justificación clínica la debe tener lógica.

## INVESTIGACIÓN Y PRÁCTICA CLÍNICA

Dada la especial importancia de las consideraciones éticas en la investigación con seres humanos, es importante diferenciarla claramente de lo que es práctica clínica (Levine, 1986). De hecho, podemos considerar que cada vez que, por ejemplo, un médico administra un fármaco a un paciente está, de algún modo, realizando un experimento. Sin embargo, el término **investigación** se utiliza para referirse al proceso sistemático y objetivo que utiliza el método científico para desarrollar o contribuir al conocimiento generalizable, mientras que el de **práctica clínica** se reserva para las actividades dirigidas exclusivamente a mejorar el estado de salud de un paciente o sujeto. Si bien es cierto que la práctica clínica se basa en una expectativa razonable de éxito, asumiendo cierto grado de incertidumbre, el hecho de que esta expectativa no esté validada no implica que se trate de una investigación, ya que para ser considerada como tal debería realizarse de acuerdo con los estándares científicos aceptados.

Existe una creciente interrelación entre la práctica asistencial y la investigación clínica. Los roles de un profesional como clínico y como investigador son muy diferentes: como clínico, su interés primario es la salud del paciente concreto al que está atendiendo, mientras que como investigador es la validez de la investigación para proporcionar un conocimiento generalizable. Estos intereses, que coexisten en el investigador clínico, pueden entrar en conflicto en determinadas ocasiones.

Cuando, como ocurre en la mayoría de los ensayos clínicos, la investigación tiene un potencial efecto terapéutico sobre los participantes, éstos tienden a creer que el estudio ha sido diseñado para su beneficio, a pesar de que el formulario de consentimiento informado explice que la finalidad es el interés científico. Este «equívoco» es aún mayor cuando el profesional que les ofrece participar en la investigación es el mismo que les trata habitualmente y la invitación se realiza en el entorno asistencial habitual.

## PRINCIPIOS ÉTICOS BÁSICOS

El Informe Belmont identifica tres principios éticos básicos: respeto por las personas o autonomía, beneficencia y justicia. Algunos autores prefieren separar el de beneficencia propiamente dicha del de no maleficencia, de manera que puede considerarse que existen dos niveles jerárquicos en estos principios: en el primer nivel se encuentran los de justicia y no maleficencia, mientras que en el segundo nivel están los de beneficencia y autonomía.

### Justicia

Este principio supone reconocer que todos los seres humanos son iguales y deben tratarse con la misma consideración y respeto, sin establecer otras diferencias entre ellos que las que redunden en beneficio de todos, y en especial de los menos favorecidos. Para ello es necesario distribuir los beneficios y las cargas de la investigación de forma equitativa.

Para cumplir este principio, la selección y el reclutamiento de los participantes deben realizarse de forma no discriminativa, asegurando que los sujetos no son seleccionados simplemente por su fácil disponibilidad, manipulabilidad o situación de dependencia que pueda hacer que se sientan obligados a participar, sino que lo son por razones directamente relacionadas con el problema de estudio. Se trata de evitar poner en situación de riesgo a determinados grupos de personas, como niños, pobres o sujetos recluidos en prisiones o instituciones cerradas, para el beneficio exclusivo de grupos más privilegiados. De hecho, la Declaración de Helsinki establece que la investigación médica en una población o comunidad con desventajas o vulnerable sólo se justifica si responde a las necesidades y prioridades de salud de esta población o comunidad y si existen posibilidades razonables de que la población o comunidad sobre la que se realiza la investigación podrá beneficiarse de sus resultados.

Además, deben contemplarse las medidas dirigidas a indemnizar a los sujetos de los posibles riesgos y perjuicios. De hecho, la normativa vigente obliga a concertar un seguro que cubra los daños y perjuicios que pudieran resultar para los sujetos como consecuencia de su participación en un ensayo clínico con un producto en fase de investigación clínica, para una nueva indicación de un medicamento ya autorizado o cuando no exista interés terapéutico para el individuo.

## No maleficencia

El principio de *no maleficencia* obliga a no infilir daño a los participantes en el estudio, ya que su protección es más importante que la búsqueda de nuevo conocimiento o el interés personal o profesional en el estudio. Por lo tanto, deben asegurarse la protección, la seguridad y el bienestar de los participantes, lo que implica, entre otras cosas, que los investigadores deben tener la formación y las calificaciones científicas apropiadas y que los centros donde se realiza el estudio deben ser adecuados.

Significa también que los riesgos para los participantes deben ser aceptables y que, si no lo son, no deben incluirse aunque los sujetos estén de acuerdo en participar.

## Beneficencia

Este principio supone procurar favorecer a los sujetos de la investigación, no exponiéndolos a daños y asegurando su bienestar. Los riesgos e incomodidades para las personas participantes deben compararse con los posibles beneficios y la importancia del conocimiento que se espera obtener, de manera que la relación sea favorable.

Implica también que la pregunta que se desea responder debe ser científicamente válida y, por lo tanto, que existe una justificación adecuada para realizar el estudio, que el diseño es correcto desde el punto de vista metodológico, ya que si no lo es, cualquier riesgo para los sujetos es innecesario, y que el número de sujetos es el adecuado, es decir, que no sea excesivo ni insuficiente.

## Autonomía

El principio de respeto por las personas o de *autonomía* se relaciona con la capacidad de una persona para decidir por ella misma. Dado que esta capacidad puede estar disminuida por diferentes motivos, como en los casos de ignorancia, inmadurez o incapacidad psíquica, cualquiera que sea su causa, o por restricciones a la libertad (como el caso de las prisiones), estos grupos *vulnerables* deben ser especialmente protegidos.

Este principio también implica garantizar la confidencialidad de la información que se recoge, asegurando la protección de los datos.

El procedimiento formal para aplicar este principio es el *consentimiento informado*, que es el otorgado por una persona que reúne los siguientes requisitos:

- Es competente o capaz legalmente para otorgar el consentimiento (la competencia legal plantea la necesidad de obtener el consentimiento a través

de un representante legal en los casos de sujetos menores de edad e incapaces).

- Ha recibido la información necesaria.
- Ha comprendido adecuadamente dicha información.
- Despues de considerar la información, ha tomado voluntariamente una decisión, libre de coacción, intimidación, persuasión, manipulación, influencia o incentivo excesivo.

Así pues, el consentimiento informado es el procedimiento que garantiza que el sujeto ha expresado voluntariamente su intención de participar en el estudio, después de haber comprendido la información que se le ha dado sobre el mismo. Desde el punto de vista ético, lo más importante no es la obtención del consentimiento, sino el proceso por el que se obtiene. Debe quedar claro que no se le pide que participe, sino que se le invita a hacerlo, y se le debe permitir reflexionar, responder a sus dudas y preguntas, y ofrecer la posibilidad de consultar con otros profesionales. Dado que la voluntariedad puede verse afectada fácilmente por la posición de autoridad e influencia del profesional y que puede existir cierta persuasión difícil de evitar en la relación médico-paciente, algunos autores sugieren que sería preferible que el consentimiento lo obtuviera una persona diferente al médico que atiende habitualmente al sujeto, para evitar el conflicto de roles entre clínico e investigador (Morin et al, 2002).

En los ensayos clínicos, la hoja de información para solicitar el consentimiento de un posible participante debe contener información sobre el objetivo del estudio, su metodología, los tratamientos que pueden serle administrados (incluyendo placebo si procede), los beneficios esperados para él o la sociedad, las molestias, incomodidades y riesgos derivados del estudio (visitas, pruebas, etc.), los posibles acontecimientos adversos, otras alternativas terapéuticas disponibles, el carácter voluntario de su participación, la garantía de que puede retirarse en cualquier momento sin perjuicios ni alteraciones de la relación médico-enfermo, los aspectos relacionados con la confidencialidad de los datos, las compensaciones económicas y los tratamientos en caso de daño o lesión, y el nombre del investigador responsable del ensayo y de resolver sus dudas, y cómo contactar con él en caso de urgencia (cuadro 3.1).

**Ejemplo 3.4.** En un análisis de 101 protocolos de ensayos clínicos aprobados en dos hospitales generales universitarios españoles, Ordovás et al (1999) evaluaron la información aportada en las hojas de información al paciente, la legibilidad formal de dichas hojas y el nivel de complejidad

**Cuadro 3.1 Elementos de información que deberían comunicarse al potencial candidato para participar en un estudio**

1. Invitación a participar en el estudio
2. Objetivos del estudio
3. Fundamentos de la selección: por qué son considerados candidatos para el estudio. Cuidado especial con la realización de pruebas destinadas exclusivamente a determinar su elegibilidad
4. Explicación de los procedimientos del estudio: duración de la participación, procedimientos que se seguirán, lugar y duración de los mismos, etc. Cuidado especial en identificar los que se derivarían exclusivamente de su participación en el estudio
5. Descripción de las incomodidades y riesgos razonablemente esperables
6. Medicación de rescate y compensaciones previstas en caso de lesión
7. Descripción de los beneficios para los sujetos y para otros que pueden esperarse razonablemente
8. Discusión de otras alternativas terapéuticas disponibles que podrían ser beneficiosas para el sujeto
9. Garantía de confidencialidad de la información
10. Consideraciones económicas: compensaciones económicas para los voluntarios sanos, costes adicionales para el sujeto, ventajas de su participación, etc.
11. Contacto para responder preguntas y resolver dudas que puedan aparecer relacionadas con la investigación
12. Posibilidad de consultar con otros profesionales
13. La renuncia a participar o continuar en el estudio no comporta castigos ni pérdidas de beneficios
14. Otros elementos, como la información a los participantes de nuevos hallazgos que puedan afectar a su participación

del vocabulario. Los apartados con mayor incumplimiento (> 30% de los casos) fueron el balance de beneficios y riesgos, la identificación y el modo de contacto con el investigador principal, la descripción de los tratamientos alternativos y la especificación de las compensaciones en caso de lesiones. Además, comprobaron que la complejidad del texto era elevada y su legibilidad deficiente, por lo que para su comprensión se requería un nivel de estudios medios-superiores o superiores en más del 90% de los casos.

## REQUISITOS ÉTICOS EN INVESTIGACIÓN

Las normas éticas son declaraciones sobre la forma en que determinadas acciones deberían (o no) realizarse en el marco de la investigación, y su finalidad es indicar cómo deben cumplirse los requisitos derivados de los principios éticos fundamentales (cuadro 3.2).

La declaración de las normas éticas en los códigos y regulaciones tiende a ser más o menos vaga, por lo que a veces pueden interpretarse de formas diferentes y puede resultar difícil saber cómo aplicarlas en situaciones concretas. En estos casos puede ser útil identificar el principio o los principios que hay detrás de la norma.

El primer requisito importante es el *valor de la pregunta de investigación*, es decir, que el conocimiento que se deriva del estudio debe tener el valor suficiente para justificar el riesgo a que se expone a los participantes. Implica que el estudio evalúa una intervención terapéutica o diagnóstica que podría conducir a mejoras en el estado de salud o bienestar; es un estudio etiológico, fisiopatológico o epidemiológico que ayudará a desarrollar dicha intervención o que contrasta una hipótesis que podría generar conocimiento relevante, aunque no tuviera una aplicación práctica inmediata. Implica también que el estudio debe estar basado en suficientes investigaciones previas de calidad, incluyendo las realizadas en laboratorios y sobre animales, y en un adecuado conocimiento de la literatura científica sobre la enfermedad o problema

**Cuadro 3.2 Requisitos éticos de una investigación (elaborados a partir de Levine [1986] y Emanuel et al [2000])**

1. Valor de la pregunta de investigación
2. Validez científica (buen diseño de la investigación)
3. Competencia de los investigadores
4. Selección de los sujetos con justicia
5. Balance favorable entre beneficios y riesgos
6. Evaluación independiente del protocolo
7. Consentimiento informado
8. Respeto por los sujetos incluidos
9. Compensación por las lesiones relacionadas con la investigación
10. Ejecución honesta del estudio
11. Comunicación puntual y precisa de los resultados

de salud, de forma que los resultados esperables justifiquen la realización del estudio.

El estudio debe tener la suficiente *validez científica* o rigor metodológico, de forma que pueda garantizarse razonablemente que conducirá a la respuesta correcta a la pregunta de investigación. Sin validez metodológica, la investigación no puede generar el conocimiento deseado, ni producir beneficio alguno ni justificar la exposición de sujetos a riesgos o molestias. Además, los participantes en un estudio asumen que de su colaboración se obtendrá un provecho, por lo que el diseño deficiente de un estudio incumple también este compromiso ético implícito.

Un aspecto que provoca controversia en la comunidad científica es el uso de placebo en los ensayos clínicos cuando existen alternativas terapéuticas efectivas disponibles. Mientras que para muchos autores no es ético utilizar un grupo que recibe placebo en estas situaciones, para otros podría ser aceptable siempre que no se perjudicara al paciente por el hecho de diferir el inicio de un tratamiento efectivo (Temple y Ellenberg, 2000). En relación con este tema, la Declaración de Helsinki establece que los posibles beneficios, riesgos, costos y eficacia de toda intervención nueva deben ser evaluados mediante su comparación con la mejor intervención probada existente. Como excepciones, considera que el uso de placebo puede ser aceptable en estudios para los que no hay una intervención probada existente, y cuando, por razones metodológicas, científicas y apremiantes, su uso es necesario para determinar la eficacia y la seguridad de una intervención, pero no implica un riesgo, efectos adversos graves o daño irreversible para los pacientes, aunque previene contra el abuso de esta opción.

El estudio debe ser llevado a cabo por *investigadores competentes*, científicamente cualificados, es decir, con la suficiente formación en metodología científica y capacidad para alcanzar los objetivos de la investigación. También deben ser clínicamente competentes, es decir, capaces de proporcionar la atención adecuada a los sujetos, por ejemplo detectando precozmente los efectos adversos que puedan aparecer o comprobando la ausencia de motivos de exclusión.

La *selección justa de los sujetos* afecta tanto a la definición de los criterios de selección como a la estrategia de reclutamiento de los participantes, de forma que solamente los objetivos científicos del estudio, y no otras consideraciones como la vulnerabilidad, el privilegio u otros factores no relacionados con el propósito de la investigación, deben guiar la determinación de los sujetos o grupos que van a ser seleccionados.

Una investigación implica el uso de fármacos, intervenciones o procedimientos sobre los que existe

incertidumbre acerca de sus *riesgos y beneficios*. Un estudio solamente está justificado cuando se han minimizado los riesgos potenciales para los sujetos y se han favorecido sus potenciales beneficios, y cuando la relación entre los beneficios para los sujetos y la sociedad y los riesgos es equilibrada o favorable a los primeros.

Es importante la *evaluación independiente* del protocolo del estudio por personas ajenas a la investigación que minimicen el posible impacto de los potenciales conflictos de intereses. Aunque hoy día la revisión y aprobación de un protocolo por los Comités Éticos de Investigación Clínica (CEIC) es un requisito legal únicamente en algunos estudios con medicamentos, cada vez hay una mayor tendencia a que estos comités evalúen cualquier tipo de investigación realizada en seres humanos.

El *consentimiento informado* es el requisito que ha recibido mayor atención. Su finalidad es asegurar que los sujetos deciden voluntariamente participar o no en un estudio, y que aceptan solamente cuando la investigación es consistente con sus valores, intereses y preferencias.

El *respeto por los sujetos* no finaliza con la firma del consentimiento, sino que debe mantenerse a lo largo de todo el estudio e incluso después, tanto si han aceptado como rechazado participar en el mismo. Implica, por ejemplo, mantener la confidencialidad de los datos recogidos sobre los sujetos candidatos, permitir que abandonen el estudio sin penalización, informar a los participantes si aparece nueva información sobre la intervención o su problema de salud que pueda ser relevante, o monitorear cuidadosamente su estado de salud durante el seguimiento.

La obligatoriedad de *compensar a los sujetos* por cualquier lesión que pueda producirse relacionada con su participación en el estudio es un requisito que cada vez se considera más importante.

El estudio debe realizarse con *honestidad*, de acuerdo con el protocolo y los mecanismos de monitorización y control de calidad suficientes para garantizar la calidad de los datos recogidos y el cumplimiento de los principios éticos.

El último requisito es la *comunicación puntual y precisa* de los resultados.

## LOS COMITÉS ÉTICOS DE INVESTIGACIÓN CLÍNICA

Uno de los componentes esenciales de la realización responsable de la investigación es su supervisión. La mayoría de los reglamentos actuales requieren la revisión y aprobación por parte de comités de ética

independientes para garantizar la protección de los seres humanos.

En España, la normativa sobre la realización de ensayos clínicos con medicamentos establece que estos estudios, antes de poder realizarse, deben contar con el informe previo del CEIC correspondiente. Su función principal es ponderar los aspectos metodológicos, éticos y legales del protocolo propuesto, así como establecer el balance de riesgos y beneficios. Para ello debe evaluar la idoneidad del protocolo y del equipo investigador, la información escrita que se proporcionará a los posibles sujetos de investigación y el tipo de consentimiento que va a obtenerse, la previsión de la compensación y el tratamiento que se ofrecerá a los sujetos en caso de lesión o muerte, así como el seguro o la indemnización para cubrir las responsabilidades, el alcance de las compensaciones que se ofrecerán a los investigadores y a los sujetos de investigación, y también debe realizar el seguimiento del ensayo clínico.

No solamente los ensayos clínicos plantean dilemas éticos, por lo que cada vez es más frecuente que se exija la aprobación por parte de un CEIC antes de poder iniciar cualquier estudio de investigación, incluso en el caso de los diseños observacionales. De hecho, muchas revistas exigen un certificado de esta aprobación como requisito necesario para la aceptación de un artículo de investigación.

## CONFLICTOS DE INTERESES EN INVESTIGACIÓN CLÍNICA

El conflicto de interés se origina cuando el juicio del profesional en relación con su interés primario (la validez del estudio en el caso del investigador) se ve influido indebidamente por un interés secundario, como el provecho económico o el afán de notoriedad. La presencia de un potencial conflicto de interés no supone por sí misma que se producirá un desenlace éticamente incorrecto, pero es evidente que incrementa su posibilidad. Además, la constatación de su existencia puede minar la confianza de las personas y de la sociedad tanto en la asistencia como en la investigación.

La investigación actual se desarrolla en un ambiente de grandes expectativas y presiones, tanto por parte de las instituciones (necesidad de producción científica para el prestigio y la obtención de fondos económicos), los promotores o patrocinadores de la investigación (presión para la obtención de resultados favorables a sus intereses), como por los propios investigadores (prestigio, respeto de los colegas o beneficios económicos). El hecho de que en nuestro país la mayor parte de la financiación de

la investigación médica corra a cargo de la industria farmacéutica, cuyo interés primario es la obtención de beneficios, propicia la aparición de múltiples conflictos de intereses (Ara, 2002):

- En el valor de la investigación, ya que se centran en estudios que pueden reportar beneficios, comparando fármacos similares y caros, buscando su aprobación y posicionamiento en el mercado.
- En el diseño de los estudios, pensado para aumentar la probabilidad de obtener resultados favorables al nuevo fármaco.
- En la inclusión de pacientes, ya que el pago de cantidades elevadas a los investigadores puede facilitar el reclutamiento indebido.
- En el análisis de los datos, a menudo inaccesibles para los investigadores, incluso con la difusión selectiva de los resultados más favorables.
- En la publicación de los resultados.
- En otras actividades relacionadas con la investigación, como los convenios económicos para actuar como conferenciantes o promotores de la utilización de fármacos.

**Ejemplo 3.5.** Kjaergard y Als-Nielsen (2002) estudiaron 159 ensayos clínicos publicados en la revista *British Medical Journal* entre enero de 1997 y junio de 2001, y observaron que las conclusiones de los autores de dichos ensayos eran más favorables a las intervenciones experimentales en los casos en que se había declarado financiación del estudio por una organización con ánimo de lucro (compañías que podían tener ganancias o pérdidas financieras en función del resultado), mientras que otros tipos de conflicto de intereses (personales, académicos, políticos, etc.) no estaban asociados con las conclusiones de los autores. Además, esta asociación no podía explicarse por la calidad metodológica, la potencia estadística, la especialidad médica ni otras variables analizadas.

**Ejemplo 3.6.** En un estudio similar, Clifford et al (2002) evaluaron 100 ensayos clínicos publicados en 1999 y 2000 en cinco revistas de medicina interna con un elevado factor de impacto, 66 de los cuales habían sido financiados en su totalidad o parcialmente por la industria farmacéutica. Aunque el porcentaje de estudios con conclusiones favorables al nuevo tratamiento evaluado era claramente superior al de estudios que favorecían al tratamiento convencional de referencia (el 67 frente al 6%), no observaron una asociación estadísticamente significativa entre la dirección de las conclusiones y la fuente de financiación de los estudios.

Por lo tanto, el hecho de que una investigación clínica se realice en colaboración con la industria farmacéutica no es por sí mismo éticamente cuestionable, sino que el profesional debe ser capaz de conjugar sus obligaciones como médico y como investigador. Para ello debe tener una formación específica en los aspectos éticos de la clínica y la investigación, así como en metodología científica, y aceptar participar solamente en estudios relacionados con su área de experiencia y que hayan sido aprobados por un CEIC, asegurándose de que la presentación y publicación de los resultados no se retrasará u obstaculizará de forma injustificada. Además, la compensación económica que recibe debe estar en consonancia con el esfuerzo que se le solicita. Lo que es cuestionable desde el punto de vista ético es no declarar o no hacer explícitos los potenciales conflictos de intereses relacionados con la investigación.

## ÉTICA Y PUBLICACIÓN CIENTÍFICA

El resultado de la investigación es conocimiento, pero para que pase a formar parte de la ciencia, este conocimiento individual debe presentarse de modo que su validez pueda juzgarse de forma independiente. No puede considerarse que una investigación ha finalizado hasta que no se ha difundido a la comunidad científica, lo que habitualmente significa su publicación en una revista y su inclusión en las bases de datos bibliográficas.

Por consiguiente, tanto los autores como los editores de las revistas médicas tienen obligaciones éticas con respecto a la publicación de los resultados de su investigación. La Declaración de Helsinki establece que los autores tienen el deber de tener a la disposición del público los resultados de su investigación en seres humanos y que son responsables de la integridad y exactitud de sus informes, y recuerda que se deben publicar tanto los resultados negativos e inconclusos como los positivos o, en caso contrario, deben estar a la disposición del público.

Así pues, los aspectos éticos de la investigación no se limitan a los relacionados con el diseño, la ejecución, el análisis y la interpretación de los resultados, sino que también incluyen la comunicación que se hace de los mismos y la forma en que se difunden a los profesionales que deben aplicarlos. Por lo tanto, la publicación científica, la realización de revisiones sistemáticas y la elaboración de guías de práctica clínica, así como la influencia que pueden tener sobre la práctica clínica, también son aspectos importantes que se deben tener en cuenta.

En el proceso de comunicación científica son cuatro los principales actores implicados:

1. Los investigadores o autores de los manuscritos, que no solamente deben haber llevado a cabo el estudio con el rigor adecuado y respetando los principios éticos, sino también haber redactado el manuscrito con claridad y honestidad. Dada la presión que existe para publicar, por motivos académicos, profesionales o económicos, pueden existir conflictos y conductas inapropiadas relacionados con aspectos como el concepto de autoría, la decisión de publicar o no un estudio, la publicación múltiple o fragmentada, la fabricación o falsificación de datos o el plagio, por ejemplo.
2. Los editores de las revistas, que deben establecer los mecanismos que garanticen la calidad de las publicaciones, y dado que son los responsables de tomar la decisión de publicar o no un trabajo, deben evitar cometer abusos desde su posición de privilegio.
3. Los asesores externos (consultores o revisores), que participan en el proceso de revisión de los manuscritos (*peer-review*) y aconsejan al comité editorial sobre la pertinencia de la publicación de los trabajos y redactan unos comentarios para los autores con la finalidad de mejorar la calidad y la presentación de los artículos. Su función es evaluar los manuscritos escrupulosamente y con imparcialidad, respetando la confidencialidad y manteniendo en todo momento el «juego limpio».
4. Los lectores, que deben ser capaces de leer críticamente los artículos que se publican, interpretarlos correctamente y valorar la aplicabilidad de los resultados a su propia práctica.

En todas las fases de la comunicación científica son múltiples los conflictos de intereses que pueden presentarse. Dado que la literatura científica es la principal fuente de evidencias que se utiliza en la toma de decisiones y en la elaboración de las guías de práctica clínica y las recomendaciones para la atención de los pacientes, el impacto que estos conflictos pueden tener sobre la práctica es muy importante.

### El proceso de revisión de manuscritos (*peer-review*)

Todo el proceso de la comunicación científica pivota sobre un mecanismo de selección de manuscritos basado en su evaluación por expertos (*peer-review*), cuyos objetivos principales son evitar la publicación de trabajos de mala calidad científica, de material no original y de trabajos que no contengan información

relevante para los lectores de la revista, así como mejorar la redacción y la presentación de los trabajos.

En la práctica, este proceso se ha convertido en un sistema de garantía de calidad de las publicaciones, con aparentes beneficios para todos los agentes implicados. Sin embargo, aun siendo reconocido como el mejor sistema disponible, es claramente insuficiente. Entre sus críticas destacan que se trata de un proceso costoso, insuficientemente contrastado, en general poco fiable (escasa concordancia entre asesores), incapaz de reconocer la investigación no original, la falsificación y fabricación de datos, el plagio, etc., poco estandarizado, abierto a todo tipo de sesgos y que produce un retraso en la publicación que para muchos es innecesario.

Por otro lado, existe una mayor probabilidad de aceptación de los estudios con resultados estadísticamente significativos, con fuentes de financiación externas y multicéntricos, mientras que la probabilidad de rechazo es mayor en los que obtienen conclusiones no acordes con la opinión de los consultores, con resultados estadísticamente no significativos y los estudios sobre intervenciones no convencionales. Como consecuencia de todo ello se produce un sesgo de publicación que puede influir sobre las conclusiones de las revisiones sistemáticas y las guías de práctica clínica basadas en la literatura disponible.

## **Conflictos de intereses y publicación científica**

El conflicto de interés puede definirse como aquel que, si se revela posteriormente, podría hacer que el lector se sintiera razonablemente decepcionado o engañado, lo que incluye también los conflictos no aparentes que pueden influir sobre el juicio del autor, los revisores o los editores.

Los conflictos pueden manifestarse de múltiples formas: falsificación o fabricación de datos, publicación selectiva de información, decisión de publicar o no un trabajo, interpretación sesgada de los resultados, etc.

**Ejemplo 3.7.** Stelfox et al (1998) identificaron los artículos publicados entre marzo de 1995 y septiembre de 1996 relacionados con la seguridad de los fármacos calcioantagonistas, clasificándolos como favorables, neutrales o críticos en relación con su uso, y solicitaron información de los autores de dichos trabajos sobre sus relaciones con la industria farmacéutica. El principal resultado fue que los autores favorables al uso de estos fármacos tenían relaciones financieras con la industria farmacéutica en un tanto por ciento superior que los que eran neutrales o críticos (100 frente al 67 y 43%, respectivamente).

**Ejemplo 3.8.** En una revisión sistemática reciente, Lexchin et al (2003) demostraron la existencia de un sesgo de manera que los estudios financiados por la industria farmacéutica tienen una mayor probabilidad de obtener resultados favorables para el fármaco que producen. Este sesgo no está relacionado con diferencias en la calidad metodológica de los estudios, sino que las explicaciones más probables son la elección de un comparador inadecuado y el sesgo de publicación.

Este problema ha llegado a ser tan importante que los editores de las principales revistas médicas publicaron conjuntamente un artículo editorial (Davidoff et al, 2001) ([www.icmje.org/sponsor.htm](http://www.icmje.org/sponsor.htm)), en el que abogaban por la declaración por parte de todos los implicados en el proceso de revisión de manuscritos y de publicación de cualquier relación que pudiera ser vista como un potencial conflicto de interés. Insisten en que ningún investigador debería participar en acuerdos que interfieran con su accesibilidad a los datos o a su capacidad para analizarlos independientemente, preparar manuscritos y publicarlos.

## **Concepto de autor**

Los investigadores están sometidos a múltiples presiones para publicar, ya sea por metas personales o exigencias institucionales, por ejemplo, lo que conduce en ocasiones a conductas inapropiadas, como hacer constar como autores a personas que no lo son (autoría regalada) o a personas inexistentes (autoría «fantasma»), o no hacer constar a todos los verdaderos autores del trabajo.

**Ejemplo 3.9.** Una encuesta realizada a los autores de 577 revisiones publicadas en la biblioteca Cochrane en el año 1999 reveló que existía una elevada prevalencia de autores que no cumplen estrictamente los requisitos: en un 39% de las revisiones había algún autor honorario, y en un 9% existía algún autor «fantasma», frecuentemente un miembro del equipo editorial Cochrane (Mowatt et al, 2002).

Según el Comité Internacional de Editores de Revistas Médicas ([www.icmje.org/](http://www.icmje.org/)), cada autor deberá haber participado suficientemente como para tomar la responsabilidad pública sobre partes del contenido del artículo. La autoría debe basarse sólo en contribuciones sustanciales a:

1. La concepción y diseño o el análisis e interpretación de los resultados.

2. La redacción y revisión crítica del contenido intelectual del artículo.
3. La aprobación de la versión final.

Concreta además que la adquisición de fondos, la recogida de datos o la supervisión general del equipo investigador por sí solos no justifican la autoría.

Debido a que cada vez los trabajos son más complejos y requieren la participación de equipos numerosos y multidisciplinarios de investigadores, resulta difícil que existan personas que cumplan estrictamente con todos los requisitos citados. Por ello existe una tendencia a solicitar que los firmantes de un trabajo especifiquen cuál ha sido su contribución al estudio (diseño, recogida de datos, análisis de los resultados, obtención de fondos, revisión intelectual del manuscrito, etc.).

## Publicación redundante, múltiple y fragmentada

Otros problemas frecuentes son la publicación de un artículo que se solapa sustancialmente con otro ya publicado, o la publicación fragmentada, en la que un estudio se divide en diferentes partes (*salami papers*) que se publican separadamente.

**Ejemplo 3.10.** Melander et al (2003) ilustran el impacto potencial de la publicación múltiple y selectiva de los estudios financiados por la industria

farmacéutica. Al analizar los estudios sometidos a la Agencia Reguladora de los Medicamentos de Suecia como base para la aprobación de cinco nuevos fármacos para la depresión mayor, se observó cómo la mitad de ellos habían contribuido al menos a dos publicaciones cada uno, y que se habían publicado con mayor frecuencia los que obtenían resultados estadísticamente significativos favorables al nuevo fármaco. También observaron que muchas publicaciones ignoraron los resultados del análisis por intención de tratar, presentando los del análisis por protocolo más favorables a los nuevos medicamentos.

El Comité Internacional de Editores de Revistas Médicas ([www.icmje.org/](http://www.icmje.org/)) sólo considera aceptable la publicación secundaria, especialmente en otro idioma, cuando se cumplen las siguientes condiciones:

1. Los autores tienen la aceptación de los editores de ambas revistas.
2. Se respeta la prioridad de la publicación primaria.
3. El artículo se dirige a un grupo diferente de lectores, por lo que una versión abreviada suele ser suficiente.
4. La versión secundaria refleja fielmente los datos e interpretación de la publicación primaria.
5. Una nota al pie de página informa de la publicación previa y la referencia.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- Angell M. The Ethics of Clinical Research in the Third World. *N Engl J Med.* 1997;337:847-9.
- Clifford TJ, Barrowman NJ, Moher D. Funding source, trial outcome and reporting quality: are they related? Results of a pilot study. *BMC Health Services Research.* 2002;2:18-23.
- Kjaergard LL, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ. *BMJ.* 2002;325:249-52.
- Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ.* 2003;326:1167-76.
- Lurie P, Wolfe SM. Unethical Trials of Interventions to Reduce Perinatal Transmission of the Human Immunodeficiency Virus in Developing Countries. *N Engl J Med.* 1997;337:853-6.
- Melander H, Ahlvist-Rastad J, Meijer G, Beermann B. Evidencie b(i)ased medicine – selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ.* 2003;326:1171-5.
- Mowatt G, Shirran L, Grimshaw JM, et al. Prevalence of honorary and ghost authorship in Cochrane reviews. *JAMA.* 2002;287:2769-71.
- Ordovás Baines JP, López Briz E, Urbieto Sanz E, Torregrosa Sánchez R, Jiménez Torres NV.
- Análisis de las hojas de información al paciente para la obtención de su consentimiento informado en ensayos clínicos. *Med Clin (Barc).* 1999;112: 90-4.
- Stelfox HT, Chua G, O'Rourke K, Detsky AS. Conflict of interest in the debate over calcium-channel antagonists. *N Engl J Med.* 1998;338: 101-6.
- Whalen CC, Johnson JL, Okwera A, Hom DL, Huebner R, Mugenyi P, et al. A trial of three regimens to prevent tuberculosis in Ugandan adults infected with the human immunodeficiency virus. Uganda-Case Western Reserve University Research Collaboration. *N Engl J Med.* 1997;337:801-8.

## BIBLIOGRAFÍA

- Ara Callizo JR. Manejo de los conflictos de intereses en la investigación médica. Visión desde la clínica. ICB digital. 2002 (30 Sept); núm. 8. Disponible en: <http://www.icbdigital.org/icbdigital/pdf/articulo/articulo8.pdf>.
- Beauchamp TL. Informed consent: its history, meaning, and present challenges. Camb Q Healthc Ethics. 2011;20:515-23.
- Bodenheimer T. Uneasy alliance. Clinical investigators and the pharmaceutical industry. N Engl J Med. 2000;342: 1539-44.
- Dal-Ré R, Tormo MJ, Pérez G, Bolívar F. Revisión ética de estudios epidemiológicos: una necesidad y una propuesta. Med Clin (Barc). 1998;111:587-91.
- Davidoff F, DeAngelis CD, Drazen JM, et al. Sponsorship, authorship, and accountability. JAMA. 2001;286:1232-4.
- Davies H, Wells F, Czarkowski M. Standards for research ethics committees: purpose, problems and the possibilities of other approaches. J Med Ethics. 2009;35:382-3.
- De Abajo FJ. La Declaración de Helsinki VI: una revisión necesaria, pero ¿suficiente? Rev Esp Salud Pública. 2001;75: 407-20.
- Ellenberg SS, Temple R. Placebo-Controlled Trials and Active-Control Trials in the Evaluation of New Treatments. Part 2: Practical Issues and Specific Cases. Ann Intern Med. 2000;133:464-70.
- Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? JAMA. 2000;283: 2701-11.
- Hossne WS, Vieira S, De Freitas CB. Committees for Ethics in Research involving human subjects. J Int Bioethique. 2008;19:131-41.
- Levine RJ. Ethics and regulation of clinical research. 2nd ed.. Baltimore: Urban & Schwarzenberg; 1986.
- Lewis JA, Jonsson B, Kreutz G, Sampaio C, van Zwieten-Boot B. Placebo-controlled trials and the Declaration of Helsinki. Lancet. 2002;359:1337-40.
- Lidz CW. Informed consent: a critical part of modern medical research. Am J Med Sci. 2011;342:273-5.
- Morin K, Rakatansky H, Riddick FA Jr, Morse LJ, O'Bannon JM 3rd, Goldrich MS, et al. Managing conflicts of interest in the conduct of clinical trials. JAMA. 2002;287:78-84.
- Neff MJ. Institutional review board consideration of chart reviews, case reports, and observational studies. Respir Care. 2008;53:1350-3.
- Newcombe JP, Kerridge IH. Assessment by human research ethics committees of potential conflicts of interest arising from pharmaceutical sponsorship of clinical research. Intern Med J. 2007;37:12-7.
- Orchard J. For debate: should observational clinical studies require ethics committee approval? J Sci Med Sport. 2008;11:239-42.
- Puri KS, Suresh KR, Gogtay NJ, Thatte UM. Declaration of Helsinki, 2008: implications for stakeholders in research. J Postgrad Med. 2009;55:131-4.
- Rosmini F. Ethics of observational studies: are specific rules useful? Epidemiol Prev. 2006;30:295-8.
- Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. Ann Intern Med. 2000;133:455-63.
- Weijer C, Dickens B, Meslin EM. Bioethics for clinicians: 10. Research ethics. CMAJ. 1997;156:1153-7.
- Weijer C, Glass KC. The ethics of placebo-controlled trials. N Engl J Med. 2002;346:382-3.

Página deliberadamente en blanco

# Parte

2

## Tipos de estudio

- |   |  |
|---|--|
| <b>4</b> Clasificación de los tipos de estudio <b>29</b>                  | <b>7</b> Estudios de cohortes <b>64</b>          |
| <b>5</b> Estudios experimentales I: el ensayo clínico aleatorio <b>33</b> | <b>8</b> Estudios de casos y controles <b>74</b> |
| <b>6</b> Estudios experimentales II: otros diseños <b>49</b>              | <b>9</b> Estudios descriptivos <b>87</b>         |
|   | <b>10</b> Revisiones sistemáticas <b>97</b>      |

Página deliberadamente en blanco

# Capítulo

4

## Clasificación de los tipos de estudio

Por *diseño* de un estudio se entienden los procedimientos, métodos y técnicas mediante los cuales los investigadores seleccionan a los pacientes, recogen datos, los analizan e interpretan los resultados.

En este capítulo se presentan los criterios de clasificación de los diseños o tipos de estudio, y en los siguientes capítulos las características, las ventajas y las limitaciones de los más empleados en la investigación clínica y epidemiológica. Conocer estos aspectos es importante para elegir el diseño más adecuado para alcanzar los objetivos fijados.

### CRITERIOS DE CLASIFICACIÓN

Las características más importantes de la arquitectura de un estudio se pueden clasificar según cuatro ejes principales:

- *Finalidad* del estudio: analítica o descriptiva.
- *Secuencia temporal*: transversal o longitudinal.
- *Control de la asignación* de los factores de estudio: experimental u observacional.
- *Inicio del estudio en relación con la cronología de los hechos*: prospectivo o retrospectivo.

#### Finalidad: descriptiva o analítica

Un estudio se considera *analítico* cuando su finalidad es evaluar una presunta relación causal entre un factor (p. ej., un agente que se sospecha que puede causar una enfermedad o un tratamiento que puede prevenir o mejorar una situación clínica) y un efecto, respuesta o resultado.

Un estudio se considera *descriptivo* cuando no busca evaluar una presunta relación causa-efecto, sino que sus datos son utilizados con finalidades

puramente descriptivas. Suele ser útil para generar hipótesis etiológicas que deberán contrastarse posteriormente con estudios analíticos.

#### Secuencia temporal: transversal o longitudinal

Se consideran *transversales* los estudios en los que los datos de cada sujeto representan esencialmente un momento del tiempo. Estos datos pueden corresponder a la presencia, ausencia o diferentes grados de una característica o enfermedad (como ocurre, por ejemplo, en los estudios de prevalencia de un problema de salud en una comunidad determinada), o bien examinar la relación entre diferentes variables en una población definida en un momento de tiempo determinado. Dado que las variables se han medido de forma simultánea, no puede establecerse la existencia de una secuencia temporal entre ellas y, por tanto, estos diseños no permiten abordar el estudio de una presunta relación causa-efecto. Así pues, los estudios transversales son por definición descriptivos.

Se consideran *longitudinales* los estudios en los que existe un lapso de tiempo entre las distintas variables que se evalúan, de forma que puede establecerse una secuencia temporal entre ellas. Pueden ser tanto descriptivos como analíticos. En estos últimos debe tenerse en cuenta, además, la dirección temporal, que puede ir de la causa hacia el desenlace (estudios experimentales y estudios de cohortes) o bien desde el desenlace hacia la causa (estudios de casos y controles).

Algunos autores consideran longitudinales sólo los estudios en los que el seguimiento de los sujetos es temporal, desde una línea basal hasta un desenlace, de modo que convierten este concepto

en sinónimo del de cohorte. Sin embargo, según la definición dada en el párrafo anterior, se considera que un estudio es longitudinal si las observaciones se refieren a dos momentos en el tiempo, aun cuando la recogida de información se haya realizado de forma simultánea. Si las distintas observaciones se han recogido en un mismo momento en el tiempo, para que el estudio pueda considerarse longitudinal se debe asumir una secuencia temporal entre ellas.

### Asignación de los factores de estudio: experimental u observacional

Se consideran *experimentales* los estudios en los que el equipo investigador asigna el factor de estudio y lo controla de forma deliberada para la realización de la investigación, según un plan preestablecido. Estos estudios se centran en una relación causa-efecto (analíticos), y en general evalúan el efecto de una o más intervenciones preventivas o terapéuticas.

Se definen como *observacionales* los estudios en los que el factor de estudio no es controlado por los investigadores, sino que éstos se limitan a observar, medir y analizar determinadas variables en los sujetos. La exposición puede venir impuesta (p. ej., el sexo o la raza), haber sido «escogida» por los propios sujetos (p. ej., el consumo de tabaco) o decidida por el profesional sanitario dentro del proceso habitual de atención sanitaria (p. ej., los actos terapéuticos ordinarios), pero no de forma deliberada en el marco de una investigación.

### Inicio del estudio en relación con la cronología de los hechos: prospectivo o retrospectivo

Los términos prospectivo y retrospectivo pueden conducir a confusión, ya que suelen aplicarse también a la dirección temporal de las observaciones, de forma que algunos autores consideran el término prospectivo como sinónimo de cohorte o incluso de longitudinal.

En este texto se consideran *prospectivos* aquellos estudios cuyo inicio es anterior a los hechos estudiados, de forma que los datos se recogen a medida que van sucediendo. Se consideran *retrospectivos* los estudios cuyo diseño es posterior a los hechos estudiados, de modo que los datos se obtienen de archivos o registros, o de lo que los sujetos o los médicos refieren. Cuando existe una combinación de ambas situaciones, los estudios se clasifican como *ambispectivos*.

En el cuadro 4.1 se presenta la clasificación de los diseños más habituales, y las figuras 4.1 y 4.2

ilustran los principales tipos de estudios que se describirán en los siguientes capítulos.

**Ejemplo 4.1.** Un estudio investiga la presunta relación causal entre el consumo de tabaco y la presencia de cardiopatía isquémica. Para ello se identifican enfermos de cardiopatía isquémica y un grupo de pacientes sin la enfermedad, y en el mismo momento se les interroga sobre su historia pasada de consumo de tabaco. Este estudio es analítico (evalúa una presunta relación causal), observacional (no se controla el factor de estudio) y retrospectivo (los hechos ya han ocurrido cuando se realiza la investigación). Aunque la información sobre el efecto y la presunta causa se recogen en un mismo momento, se asume que los datos sobre el consumo de tabaco se refieren a un momento del tiempo anterior a la aparición de la enfermedad, por lo que este estudio puede clasificarse como longitudinal (dado que la dirección es de efecto a causa, correspondería a un estudio de casos y controles).

Cuadro 4.1 Clasificación de los tipos de diseño

#### Estudios descriptivos transversales

- Estudios de prevalencia
- Series de casos transversales
- Evaluación de pruebas diagnósticas
- Estudios de concordancia
- Estudios de asociación cruzada
- Otros estudios transversales descriptivos

#### Estudios descriptivos longitudinales

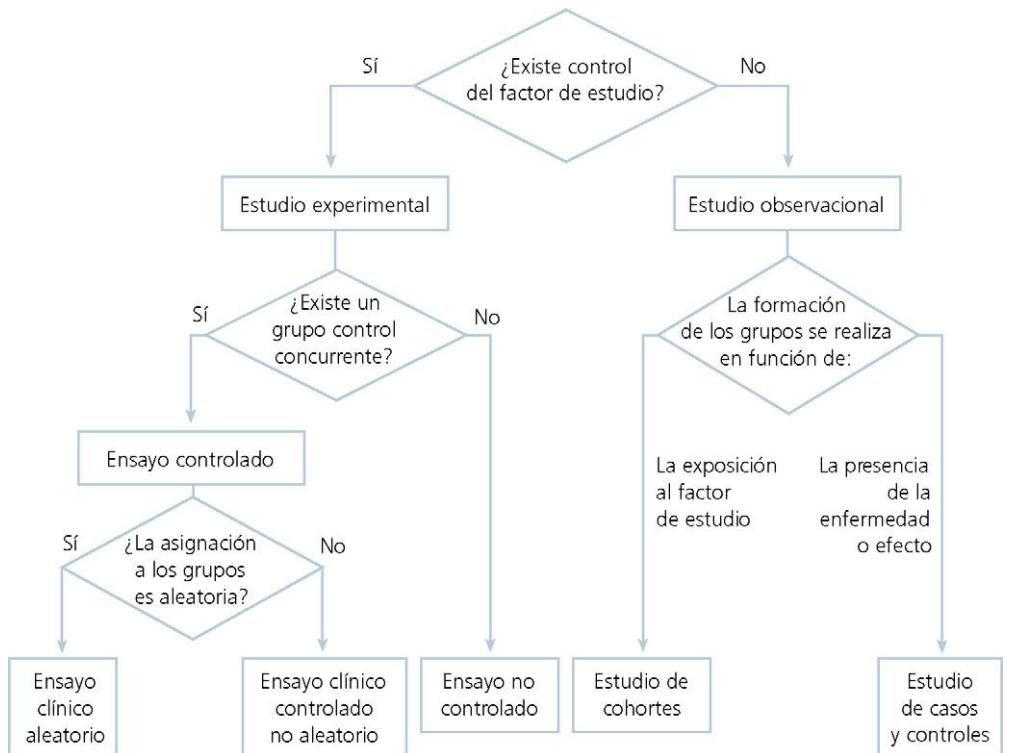
- Estudios de incidencia
- Descripción de los efectos de una intervención no deliberada
- Descripción de la historia natural

#### Estudios analíticos observacionales

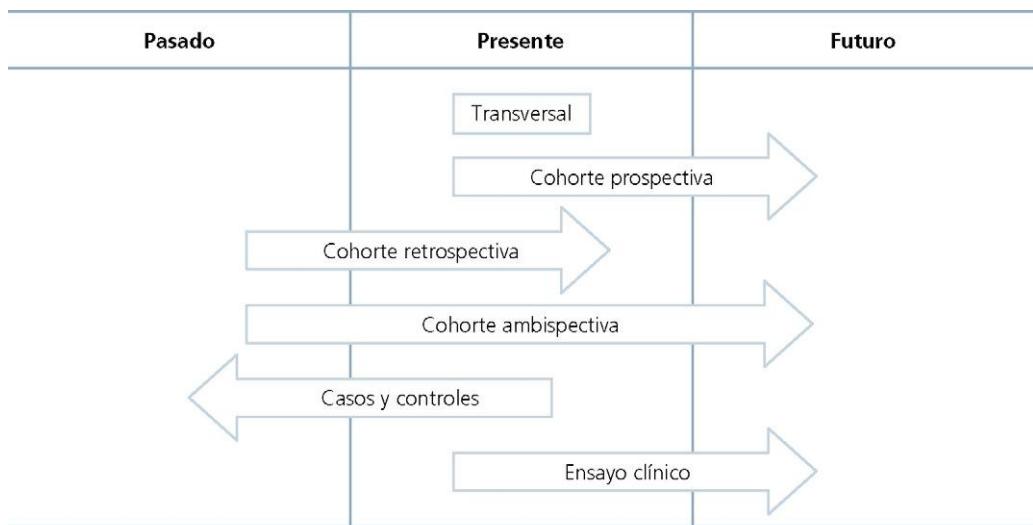
- Dirección causa-efecto: estudios de cohortes:
  - Prospectivos
  - Retrospectivos
  - Ambispectivos
- Dirección efecto-causa: estudios de casos y controles
- Estudios híbridos

#### Estudios analíticos experimentales

- Ensayos controlados:
  - Ensayos clínicos en paralelo
  - Ensayos clínicos cruzados
  - Ensayos comunitarios
- Ensayos no controlados:
  - Ensayos sin grupo control
  - Ensayos con control externo



**Figura 4.1** Algoritmo de clasificación de los estudios analíticos.



La flecha implica longitudinalidad y su sentido indica la dirección temporal.

**Figura 4.2** Principales tipos de estudios.

Supongamos que en este mismo estudio se determinan las cifras de colesterolemia en las primeras horas del infarto agudo de miocardio. Dado que es conocido que los cambios metabólicos que acompañan al infarto pueden alterar las cifras de colesterolemia, no podría asumirse que estos valores reflejan la situación previa de los sujetos, por lo que en este caso el estudio se consideraría transversal, ya que la presencia de la enfermedad y la colesterolemia corresponden a un mismo momento. Para poder considerarlo longitudinal, debería disponerse de información sobre su historia previa de colesterolemia.

**Ejemplo 4.2.** Un estudio desea evaluar la utilidad de un marcador biológico en el diagnóstico de un determinado tipo de cáncer. Para ello se selecciona una muestra de pacientes con dicho tipo de cáncer y otra de sujetos sin él, y se miden en todos ellos los valores del marcador biológico, evaluando si es útil para diferenciar ambos grupos. Se trata de un diseño descriptivo (no evalúa una presunta relación causal), transversal (ya que la identificación de la existencia del cáncer y la medición del marcador biológico se refieren al mismo tiempo) y observacional (no se controla el factor de estudio).

El mismo objetivo de investigación se hubiera podido abordar con un diseño longitudinal: medir en una muestra de sujetos el marcador biológico, realizar un seguimiento temporal para observar cuántos de ellos desarrollan el cáncer, y evaluar si los valores del marcador pueden ser predictivos de su aparición.

**Ejemplo 4.3.** Supongamos que se selecciona una muestra aleatoria de una población, y en todos los sujetos se miden las cifras de presión arterial (PA) con la finalidad de conocer cuántos tienen hipertensión arterial (HTA). Se trata de un estudio descriptivo, observacional y transversal, que estima la prevalencia de HTA (estudio de prevalencia).

Supongamos que en la misma muestra de sujetos se miden también las cifras de colesterol sérico y se correlacionan con las de PA. El diseño tiene las mismas características, por lo que, al ser transversal, no puede establecer una relación causal entre ambas variables, sino que se limita a evaluar si existe asociación entre ellas en una muestra de sujetos en un momento determinado (estudio de asociación cruzada).

**Ejemplo 4.4.** Un investigador selecciona una muestra de 30 sujetos hipertensos a los que administra un tratamiento con una pauta preestablecida, y realiza su seguimiento durante 12 meses para determinar cuántos reducen sus cifras de PA. El estudio trata de evaluar una presunta relación causal entre el tratamiento y la reducción de la PA (analítico), es experimental (ya que el investigador controla de forma deliberada el tratamiento para la realización del estudio), longitudinal y prospectivo. En este estudio no existe un grupo control que reciba un placebo u otro tratamiento, por lo que corresponde a un ensayo sin grupo control (estudio antes-después).

Supongamos que otro investigador ha seleccionado las historias clínicas de 30 pacientes hipertensos que han recibido el mismo tratamiento, y ha recogido una serie de variables, entre ellas la PA al inicio del tratamiento y al año de éste. Dado que la intervención no se ha administrado de manera deliberada para el estudio, no corresponde a un diseño experimental. De hecho, es la descripción de una cohorte de sujetos hipertensos que reciben un tratamiento dentro de la práctica asistencial, que sólo permite conocer la evolución de las cifras de PA y que, por tanto, no puede considerarse analítico, ya que no permite evaluar la relación causal entre el tratamiento y las cifras de PA. Además, dado que los datos proceden de registros clínicos, se clasificaría como retrospectivo.

## BIBLIOGRAFÍA

Fletcher RH, Fletcher SW.  
Clinical Epidemiology. The  
essentials. Philadelphia:  
Lippincott Williams & Wilkins;  
2005.

Hulley SB, Cummings SR,  
Brownwe WS, Grady DG.  
Designing etnical research.  
Philadelphia: Lippincott  
Williams & Wilkins; 2007.

Levin KA. Study design I.  
Evidence-based. Dentistry.  
2005;6:78-9.

# Capítulo

# 5

## Estudios experimentales I: el ensayo clínico aleatorio

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

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

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

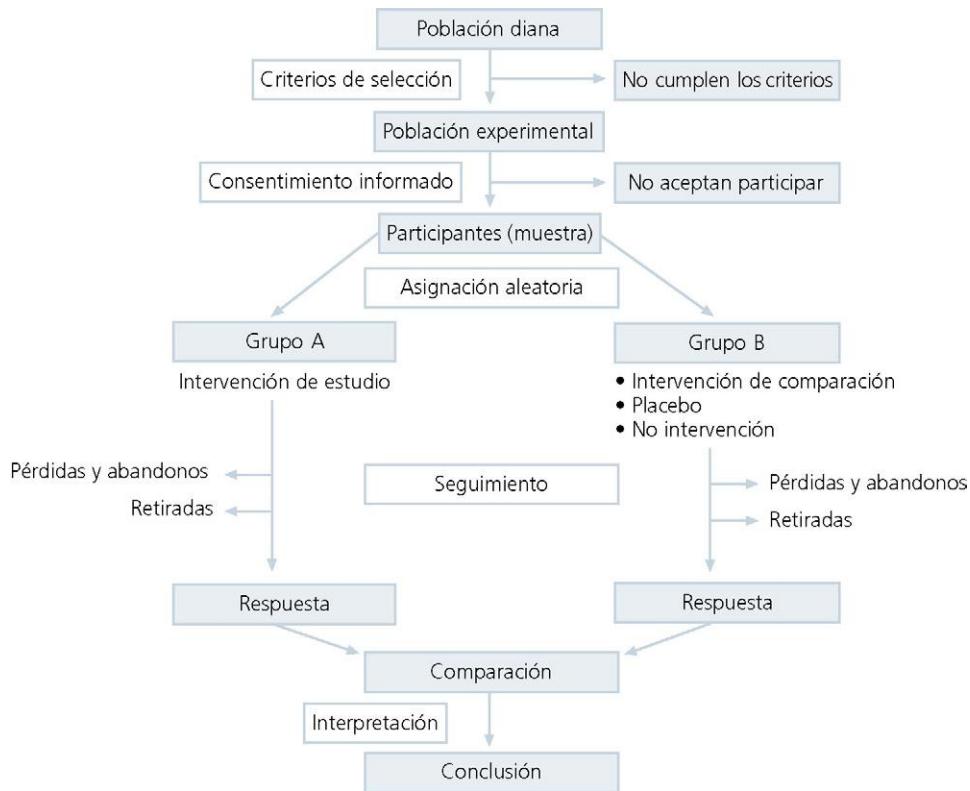
El diseño experimental más importante es el ensayo clínico aleatorio (ECA). Aunque existen diversas formas de diseñarlo, habitualmente esta denominación se utiliza para referirse al diseño

paralelo, que es el más conocido y utilizado, y cuyas características generales se describen en este capítulo. Las otras formas de ensayo clínico y del resto de diseños experimentales se abordan en el capítulo siguiente.

### ENSAYO CLÍNICO ALEATORIO

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

El ECA se considera el mejor diseño disponible para evaluar la eficacia de una intervención sanitaria, ya que es el que proporciona la evidencia de mayor calidad acerca de la existencia de una relación causa-efecto entre dicha intervención y la respuesta observada. Por este motivo, cada vez con más frecuencia, las recomendaciones terapéuticas y las guías de práctica clínica se basan en las pruebas proporcionadas por este tipo de estudios. Ello ha conducido a que el número de ECA esté aumentando rápidamente y a que cada vez se exija un mayor rigor metodológico en su diseño, ejecución y análisis. Además, ha llevado al desarrollo de instrumentos para la valoración de su calidad metodológica, guías para su lectura rápida y recomendaciones sobre su publicación, como la propuesta CONSORT. Las principales



**Figura 5.1** Estructura básica de un ensayo clínico aleatorio paralelo.

ventajas e inconvenientes de un ECA se resumen en el cuadro 5.1.

Las autoridades sanitarias obligan a la realización de ensayos clínicos para demostrar la eficacia y seguridad de un nuevo fármaco antes de su comercialización, o bien de una nueva forma terapéutica, una nueva indicación o su uso en condiciones diferentes de las autorizadas. Pero el ECA es también el diseño preferible para comparar en la práctica diferentes pautas terapéuticas autorizadas, o incluso intervenciones sanitarias de cualquier tipo, como tratamientos no farmacológicos, consejos sanitarios, recomendaciones dietéticas, detección precoz de enfermedades, estrategias de formación médica continuada, o incluso modelos organizativos de las consultas o los centros sanitarios.

A continuación se describen las características generales del diseño de este tipo de estudios, dejando la discusión más detallada de los diferentes elementos del protocolo para los capítulos correspondientes.

## INTERVENCIONES QUE SE COMPARAN

Uno de los aspectos clave del diseño de un ECA es la selección de la intervención que se va a utilizar como referencia en la comparación. Debe respetarse el denominado principio de incertidumbre (*equipoise*), según el cual un ECA solamente debe realizarse si, a la luz de los conocimientos disponibles, ambas intervenciones que se comparan pueden considerarse razonablemente alternativas terapéuticas para los pacientes, ya que existen dudas acerca de si una de ellas es superior a la otra. La comparación frente a intervenciones que se sabe que son inferiores, además de ser éticamente inaceptable, conduce a la obtención de resultados favorables a la intervención de estudio, cuya publicación introduce un sesgo en las evidencias disponibles sobre la eficacia de los tratamientos, con las repercusiones que este hecho puede tener sobre las decisiones y las recomendaciones terapéuticas.

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

**Ventajas**

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

**Desventajas**

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

concluyen que el sesgo observado en la publicación de la investigación financiada por la industria podría ser la consecuencia de violaciones del principio de incertidumbre.

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

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

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

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

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

Cuando existe una opción terapéutica aceptada como eficaz en la situación clínica de interés, ésta debería ser la alternativa de comparación. La

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

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

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

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

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

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

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

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

## DEFINICIÓN DE LA VARIABLE DE RESPUESTA

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

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

Son variables relevantes para los pacientes la calidad de vida o la recuperación de una capacidad funcional determinada, por ejemplo. Muchos ensayos se centran exclusivamente en los beneficios potenciales de un tratamiento, pero no miden sus efectos sobre otras variables importantes o incluso algunos efectos secundarios y adversos. La medición de variables que incluyan la perspectiva del paciente será importante para establecer un adecuado balance entre beneficios y riesgos.

También son importantes las variables clínicas, como la supervivencia en el cáncer, las fracturas vertebrales en la osteoporosis, los síntomas o medidas funcionales como la frecuencia de la angina o la tolerancia al ejercicio, o la aparición de sucesos graves e irreversibles, como el infarto agudo de miocardio (IAM) o el accidente cerebrovascular (ACV), por ejemplo. Cada vez se utilizan más las llamadas variables compuestas, que combinan diferentes resultados clínicos en una única variable, lo que permite aumentar el número de desenlaces que se observan.

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

## SELECCIÓN DE LA POBLACIÓN

La definición del objetivo del ensayo hace referencia a la población diana a la que se desea poder extraer los resultados. Sin embargo, el estudio se lleva a cabo sobre una población definida por unos criterios de selección especificados a priori (población experimental), de la que proceden los sujetos que finalmente participarán en el ensayo.

Los criterios de selección identifican una población en la que, a la luz de los conocimientos actuales, las intervenciones que se comparan podrían estar igualmente indicadas y, por tanto, potencialmente

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

## Consentimiento informado

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

Antes de otorgar su consentimiento, un candidato debe recibir información, tanto oral como escrita, sobre lo que supone la experiencia y las posibles consecuencias que pueden derivarse de su participación. Esta información debe incluir como mínimo: *a)* el objetivo del estudio; *b)* las características básicas del diseño (asignación aleatoria, técnicas de enmascaramiento, uso de placebo, etc.); *c)* los posibles efectos secundarios; *d)* los posibles beneficios; *e)* que el paciente tiene el derecho de abandonar el estudio en cualquier momento, y *f)* que tiene el derecho de formular cuantas preguntas deseé relacionadas con la investigación. Todos estos puntos deben explicarse en términos comprensibles para el paciente, evitando el lenguaje académico o científico.

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

## Período de preinclusión

Algunos ECA utilizan un período previo a la inclusión de los sujetos y a su asignación a los grupos

de estudio (*run-in phase*), durante el cual se realiza el seguimiento de los pacientes con diferentes finalidades.

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

**Ejemplo 5.5.** En el American Physicians' Health Study (Hennekens y Eberlein, 1985), diseñado para evaluar si 325 mg de ácido acetilsalicílico (AAS) administrado a días alternos reduce la mortalidad cardiovascular en pacientes asintomáticos, los 33.223 sujetos que cumplían los criterios de selección establecidos fueron sometidos a un período de preinclusión de 18 semanas durante el cual todos ellos recibieron AAS, de manera que solamente fueron incluidos en el estudio los 22.071 sujetos que cumplieron con la intervención durante dicho período y no presentaron efectos secundarios. Estos participantes fueron asignados aleatoriamente a dos grupos (AAS y placebo). Como consecuencia de este proceso de selección tan estricto, a los 57 meses de seguimiento, prácticamente el 90% de los participantes todavía cumplía con la intervención asignada.

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

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

**Ejemplo 5.6.** Un ensayo clínico tenía por objetivo comparar la eficacia de dos alternativas terapéuticas en pacientes que no hubieran respondido a una dosis de 5 mg/día de amlodipino (Marín-Iranzo et al, 2005). Para ello, se diseñó un período de preinclusión durante el cual los pacientes hipertensos recibieron placebo durante 2 semanas y después 5 mg/día de amlodipino durante otras 4 semanas. De los 328 sujetos incluidos inicialmente, 83 fueron excluidos en este período porque se controlaron sus cifras de presión arterial diastólica o presentaron acontecimientos adversos, de manera que finalmen-

te tan sólo 245 pacientes fueron asignados aleatoriamente a los dos grupos que se comparaban: un grupo en el que la dosis de amlodipino se aumentó a 10 mg/día y otro grupo al que se administró una combinación de enalapril y nitrendipino. Así pues, la principal utilidad del período de preinclusión en este estudio fue seleccionar a los sujetos no respondedores.

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

Para poder interpretar correctamente los hallazgos de los estudios con período de preinclusión, es importante señalar las diferencias entre las características iniciales de los sujetos que han sido excluidos durante dicho período y los que son asignados a los grupos de estudio.

## ASIGNACIÓN ALEATORIA

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

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

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

Dado que la asignación aleatoria es la clave de un ECA, debe realizarse correctamente, de modo que ni las preferencias del médico ni las del paciente influyan en la decisión del grupo al que éste es asignado. Por ello, es esencial que se produzca después de que el paciente haya sido incluido en el estudio y que la secuencia de asignación esté oculta para el investigador, de manera que el conocimiento de los grupos no influya en su decisión de incluirlo en uno u otro. Generalmente, la asignación se realiza tras completar una serie de pruebas (cuestionarios, examen físico, exploraciones complementarias, etc.) para comprobar si los pacientes cumplen o no los criterios de selección. En caso afirmativo, el paciente deberá dar su consentimiento antes de ser incluido. Una vez cumplidos estos dos requisitos se procederá a asignar aleatoriamente a los pacientes, pero no antes. De este modo, la decisión de incluir o no a un sujeto no estará influida por el conocimiento del tratamiento que recibirá.

## TÉCNICAS DE ENMASCARAMIENTO

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

### Tipos de enmascaramiento

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

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

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

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

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

### Ventajas del enmascaramiento

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

Estas preferencias también se dan en los pacientes. El investigador contagiará su entusiasmo o su pesimismo sobre la nueva intervención a los participantes en el estudio, lo que influirá sobre la respuesta. Si el paciente está convencido de la eficacia de un nuevo tratamiento, juzgará de forma más benéfica sus efectos secundarios y tenderá a sentirse mejor, aunque sea como resultado de la autosugestión.

Por ello, siempre que sea posible, y especialmente cuando en la medición de la respuesta pueda influir el investigador o el paciente, es importante el uso de técnicas de enmascaramiento.

### Limitaciones del enmascaramiento

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

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

De estos resultados surgen dos interrogantes: ¿ha podido influir en los resultados globales el elevado tanto por ciento de aciertos del grupo al que pertenecía cada sujeto? ¿Cómo se puede mejorar el método doble ciego? En el ejemplo, como comentan los autores del estudio, la mayoría de los efectos adversos aparecieron tarde y los análisis de microalbuminuria se conocieron a los 6 y 12 meses de la asignación aleatoria, por lo que difícilmente pudieron condicionar los resultados.

El método de doble ciego puede perfeccionarse mediante la ocultación de datos no relevantes para el cuidado del paciente. En el ejemplo anterior, los valores de microalbuminuria podían haberse comunicado al equipo investigador cuando el estudio hubiese finalizado. El doble ciego también puede perfeccionarse con la participación de dos observadores en la evaluación y el desarrollo del estudio, de forma que uno de ellos no pueda tener acceso a información reveladora del fármaco administrado antes de conocer el código de asignación aleatoria.

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

**Ejemplo 5.8.** Supongamos un estudio cuyo objetivo es evaluar si el consejo médico efectuado de forma estructurada es útil en el tratamiento de la depresión posparto, en el que la intervención consiste en varias visitas al domicilio de las pacientes a cargo de profesionales sanitarios previamente entrenados. Las pacientes asignadas al grupo de estudio no podrían estar «ciegas» a la intervención.

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

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

**Ejemplo 5.9.** En el *Multiple Risk Factor Intervention Trial (MRFIT)* (1982), individuos de mediana edad con un riesgo elevado de padecer cardiopatía isquémica fueron asignados de forma aleatoria con el fin de evaluar si la intervención simultánea sobre el tabaco, la dieta y la presión arterial prevenía la enfermedad. En el grupo experimental, el control de la hipercolesterolemia y de la hipertensión arterial, así como los consejos dietético y antitabaco, estaba a cargo de un grupo de profesionales altamente especializados, mientras que en el otro grupo era llevado a cabo por sus respectivos médicos de cabecera. Los investigadores esperaban que el grupo experimental conseguiría reducir sus cifras de colesterol en un 10% si éste era mayor de 220 mg/dl, y un 10% en la presión diastólica inicial en el caso de que fuera mayor o igual a 95 mmHg, mientras que en el grupo control, tanto el colesterol como la presión arterial permanecerían estables. También se esperaba que el número de cigarrillos fumados disminuiría en ambos grupos, aunque de manera más pronunciada en el experimental. Sin embargo, las predicciones no se cumplieron, y en el grupo control los valores de los distintos factores de riesgo

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

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

## SEGUIMIENTO DE LOS SUJETOS

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

**Ejemplo 5.10.** Consideremos un estudio cuyo objetivo era analizar si los métodos de citación sistemáticos para la práctica de un examen preventivo, en concreto una citología de cérvix, eran más efectivos que uno no sistemático, y cuál de los dos métodos sistemáticos era el más efectivo (Pierce et al, 1989). La metodología empleada fue la de ECA, y su duración fue de un año. Las pacientes procedían de un centro de salud, y la intervención consistió, para el primer grupo, en el envío de una carta en la que se las invitaba a acudir al centro para el examen; las participantes del segundo grupo poseían historias

clínicas marcadas con una etiqueta, con el fin de que el médico les recordara la conveniencia de practicar el examen; en el tercero no se realizó ninguna intervención especial. La variable de respuesta analizada fue la práctica del examen preventivo durante el año que duró el estudio.

Los principales resultados fueron los siguientes: la citología se había practicado al 32% de las mujeres del primer grupo, al 27% de las del segundo grupo y al 15% de las del tercer grupo. La mayoría de las personas del primer grupo que acudieron al centro para la realización del examen preventivo, lo hicieron durante las primeras 8 semanas del estudio. En el segundo grupo, la citología se fue practicando durante todo el año, a medida que las participantes acudían al centro. Sin embargo, sólo el 73% de las mujeres asignadas a este grupo consultó con sus respectivos médicos durante el año de duración, por lo que más de una cuarta parte no tuvo la oportunidad de ser visitada. Por el contrario, todas las mujeres del grupo que recibió la carta fueron visitadas y tuvieron la oportunidad de que se les practicara la citología de cérvix. El tiempo de seguimiento, como comentan los propios autores, no fue el más adecuado para el segundo y el tercer grupos. Si se hubiera prolongado por espacio de 2 o 3 años, el porcentaje de respuestas quizás habría aumentado en ellos.

El seguimiento debe adecuarse a cada problema concreto y ser lo suficientemente prolongado como para permitir que se presente la respuesta esperada. En algunos casos será de pocas semanas, como sucede con las infecciones urinarias. En otros, se alargará durante años, en especial cuando se evalúan medidas de prevención primaria. Cuanto menor sea el tiempo de seguimiento, más fácil será mantener el contacto con los participantes, así como el interés y la motivación de los investigadores por el estudio, y por consiguiente, la probabilidad de pérdidas será menor.

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

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

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

de los motivos y del hecho de que su respuesta a la intervención sea distinta de la observada en los individuos que finalizan el estudio.

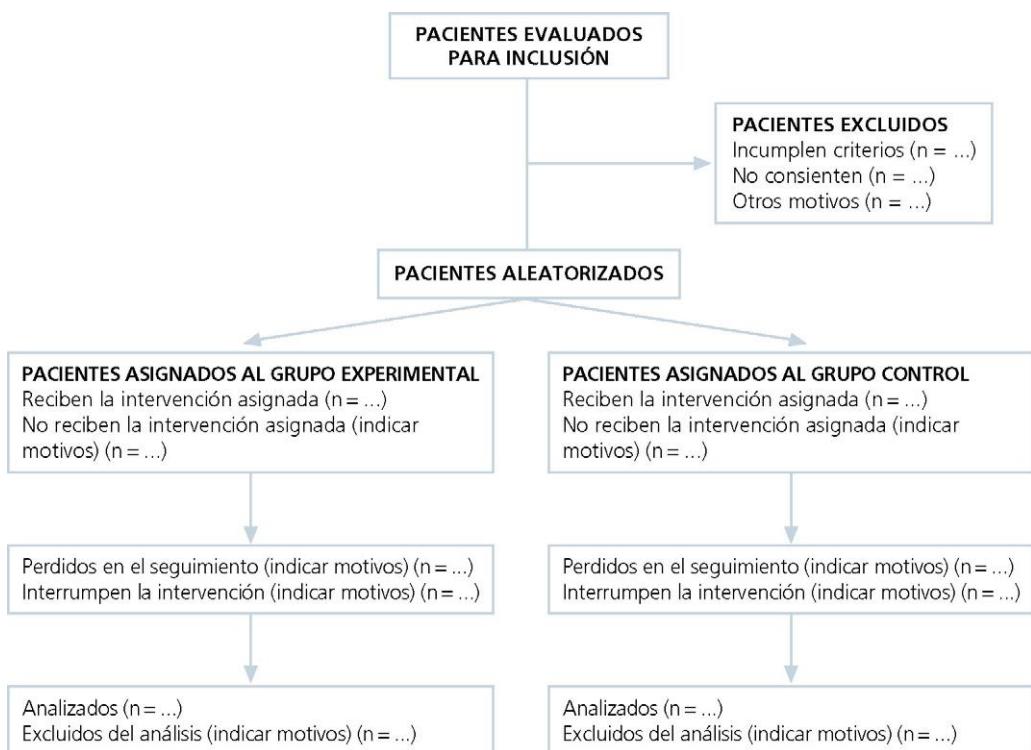
**Ejemplo 5.11.** Supongamos un ECA en el que se incluyen 200 sujetos, asignados a dos grupos de 100 individuos cada uno. Finalizan 80 en cada grupo, de los que 40 han presentado un resultado positivo, por lo que la eficacia ha sido del 50% en ambos grupos entre los sujetos que han finalizado el estudio. Sin embargo, este resultado puede estar sesgado por las pérdidas que se han producido. Supongamos que los 20 sujetos que se pierden en el grupo A lo han hecho por motivos no relacionados con el estudio, de forma que 10 de ellos han presentado un resultado positivo. Por tanto, la eficacia en los 100 pacientes del grupo A es de  $(40 + 10)/100 = 50\%$ . Por el contrario, las pérdidas en el grupo B no se han producido al azar, sino que se han debido a los efectos secundarios de la intervención, de forma que en ninguno de ellos se ha observado un resultado positivo. Por consiguiente, la eficacia en los 100 individuos del grupo B es de  $(40 + 0)/100 = 40\%$ . Este

ejemplo ilustra cómo, aunque se haya producido un mismo número de pérdidas en ambos grupos, si sus motivos son diferentes, pueden introducir un sesgo en los resultados.

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

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

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



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

## ESTRATEGIA DE ANÁLISIS

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

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

Por un lado, la exclusión de sujetos u observaciones del análisis disminuye la potencia estadística, ya que el número de individuos que se tiene en cuenta es inferior al previsto en el inicio, por lo que existirá una menor capacidad para detectar la diferencia o asociación de interés. Pero lo más importante es que, si el porcentaje de pérdidas o abandonos, o los motivos por los que determinados sujetos no son considerados en el análisis, son diferentes entre los grupos, estas situaciones especiales pueden alterar la comparabilidad. En general, las pérdidas y las violaciones del protocolo, o por lo menos algunas de ellas, no se producen al azar, sino que pueden depender de los tratamientos administrados. Por tanto, la decisión de excluir determinados sujetos del análisis puede introducir un sesgo y comprometer la validez interna del estudio.

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

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

Las exclusiones que se producen antes de la asignación aleatoria no originan ningún sesgo en la comparación de los datos, es decir, no afectan a la validez interna. Sin embargo, es importante conocer

su número y sus características para poder valorar su impacto sobre la generalización de los resultados.

## FINALIZACIÓN ANTICIPADA

A veces es útil incluir una regla para finalizar el estudio antes de lo previsto, cuando el resultado ya es suficientemente claro. En estas situaciones no es ético que un grupo de sujetos siga recibiendo el tratamiento que se ha mostrado menos eficaz o más perjudicial. Este tipo de reglas se suelen incorporar en la mayoría de los estudios con un elevado número de pacientes y que comportan un seguimiento de varios años.

Para conseguirlo, los resultados del estudio han de ser monitorizados y se realizan análisis intermedios en momentos prefijados para considerar si es probable que la continuación del estudio produzca respuestas más concluyentes o amplias. La realización de múltiples comparaciones debe tenerse en cuenta en el diseño, tanto al determinar el número de sujetos necesarios como al fijar el grado de significación estadística.

Se recomienda que la decisión de finalizar anticipadamente un ensayo recaiga sobre un grupo de expertos que no esté implicado en el diseño ni en el seguimiento del estudio. Tomar esta decisión como consecuencia de la demostración de efectos beneficiosos debe exigir una evidencia muy convincente. Interrumpir un ensayo cuando esta evidencia convincente estaba disponible precozmente podría considerarse poco ético, sobre todo cuando un fármaco está probándose frente a un placebo. La decisión de finalizar un ensayo clínico a causa de unos resultados negativos es todavía más compleja y se acepta que pueda exigirse disponer de evidencias menos concluyentes. En el cuadro 5.3 se resumen los motivos de interrupción prematura de un ensayo clínico.

**Ejemplo 5.12.** El American Physicians' Health Study es un ensayo clínico doble ciego controlado con placebo, diseñado para evaluar si 325 mg de ácido acetilsalicílico (AAS) administrado a días alternos reduce la mortalidad cardiovascular en pacientes asintomáticos. Algunas de sus características se han comentado en el ejemplo 5.5. El estudio se suspendió de forma anticipada tras un promedio de 60,2 meses de seguimiento de los 22.071 participantes, al observarse una importante reducción relativa del riesgo de IAM del 44% en el grupo de AAS en relación con el de placebo (riesgo relativo = 0,56; IC 95%: 0,45-0,70;  $p < 0,00001$ ). Se consideró que la evidencia de que el AAS reducía el riesgo de IAM era suficientemente concluyente para justificar la finalización anticipada del ensayo, a pesar de que las pruebas sobre la incidencia

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

#### Por datos generados por el ensayo clínico

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

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

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

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

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

de ACV y la mortalidad cardiovascular total eran todavía insuficientes, dado el escaso número de eventos observados (Steering Committee of the Physicians' Health Study Research Group, 1989).

## ESTUDIOS DE EQUIVALENCIA Y NO INFERIORIDAD

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

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

El diseño de los estudios de equivalencia y no inferioridad presenta algunas diferencias importantes en relación con los de superioridad, y debe ser especialmente riguroso. En primer lugar, debe asegurarse la imparcialidad de la comparación, es decir, que las condiciones del ensayo no favorecen a ninguna de las intervenciones por encima de la otra (dosis y pautas óptimas para ambas, evitar subgrupos de pacientes que podrían manifestar de forma distinta los efectos de los tratamientos, variables y tiempo de valoración adecuados). Dado que el hecho de que dos intervenciones sean equivalentes no implica necesariamente que ambas sean eficaces, es importante que el ensayo se diseñe con criterios similares a los de los estudios en los que la intervención de comparación ya ha demostrado suficientemente su eficacia.

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

Uno de los aspectos clave de este tipo de estudios es, precisamente, establecer el margen de equivalencia, que corresponde a la máxima diferencia entre ambos tratamientos que se considera clínicamente aceptable por su escasa relevancia clínica. Este valor se conoce como delta, y la finalidad del estudio es precisamente descartar estadísticamente su existencia. Se trata de demostrar que la nueva intervención es lo suficientemente similar a la de referencia como para considerarlas equivalentes desde el punto de vista clínico.

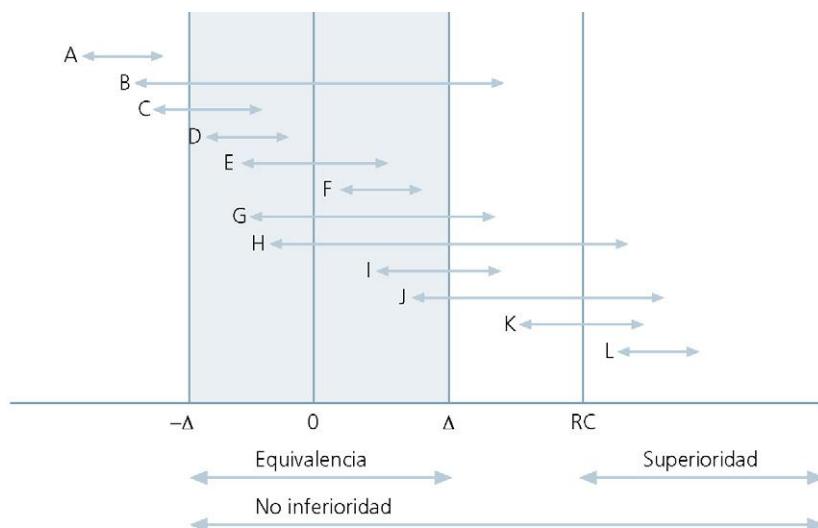
**Ejemplo 5.13.** Birtwhistle et al (2004), ante la escasa evidencia sobre la periodicidad adecuada de las visitas de seguimiento de los pacientes hipertensos, compararon el control de la presión arterial, la satisfacción y el cumplimiento del tratamiento por parte de estos pacientes, según si las visitas se realizaban cada 3 o cada 6 meses. Para ello, diseñaron un ensayo clínico en el que participaron 50 médicos

de familia de Canadá. Los 609 pacientes incluidos fueron asignados a dos grupos, y de cada uno se realizó el seguimiento con una de las estrategias que se comparaban. La asignación fue aleatoria, estratificada por médico y por bloques de 8 pacientes para asegurar que cada médico tenía un número de pacientes similar en cada grupo. El diseño fue planteado como un estudio de equivalencia, con la hipótesis de que la verdadera diferencia en el porcentaje de pacientes controlados entre los grupos sería inferior al 10%. A los 3 años de seguimiento, el porcentaje de pacientes controlados, el grado de satisfacción y el cumplimiento del tratamiento fueron similares en ambos grupos, sin diferencias relevantes ni estadísticamente significativas. Los autores concluyeron que la estrategia de seguimiento de visitas cada 6 meses es suficiente, ya que es equivalente a la de cada 3 meses.

El análisis de los resultados de estos estudios se basa en el cálculo del intervalo de confianza de la diferencia en la respuesta observada con ambas

intervenciones (fig. 5.3). En los ensayos de equivalencia, si todo el intervalo cae dentro del rango de equivalencia definido por el valor delta prefijado, puede concluirse que las intervenciones son equivalentes, con una escasa probabilidad de error, mientras que si cae fuera de dicho rango, se concluye que no son equivalentes. En el resto de situaciones, los resultados del estudio no son concluyentes. En los ensayos de no inferioridad, el enfoque es similar, pero se tiene en cuenta tan sólo el margen de no inferioridad definido por el valor delta.

Por otro lado, en los estudios de superioridad, la estrategia de análisis por intención de tratar es la más conservadora, ya que la inclusión en el análisis de los sujetos que no han cumplido el protocolo tiende a hacer más similares los resultados de ambos grupos. Sin embargo, en los estudios de equivalencia y no inferioridad, de lo que se trata es precisamente de evitar cualquier influencia que pueda hacer que los grupos parezcan más similares de lo que son en realidad, por lo que el análisis por



$\Delta$ : margen de equivalencia.

RC: mínima diferencia de relevancia clínica.

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

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

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

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

intención de tratar no es el más adecuado. En cambio, el análisis por protocolo tiende precisamente a resaltar cualquier diferencia entre los grupos más que a disminuirla, por lo que es preferible en estos estudios, aunque en algunas ocasiones, dependiendo de los motivos por los que se han producido las violaciones del protocolo, también puede distorsionar los resultados hacia la conclusión de no diferencia. Por ello, es preferible realizar ambas estrategias de análisis y esperar que coincidan, así como recoger información detallada de todos los sujetos incluidos, lo que permitirá una mayor flexibilidad en el análisis y proporcionará una base más fuerte para la conclusión del estudio.

## ENSAYOS PRAGMÁTICOS Y ENSAYOS EXPLICATIVOS

Como se ha podido comprobar, al diseñar un ECA deben tomarse múltiples decisiones, como cuál es el comparador, qué criterios de selección deben utilizarse, cuál es la pauta de seguimiento más adecuada, etc. En general, en el diseño pueden adoptarse dos posturas contrapuestas: explicativa y pragmática (tabla 5.1).

La *actitud explicativa* consiste en establecer criterios de selección muy estrictos, que definan una población muy homogénea, con escasa variabilidad, buena cumplidora, en la que sea más fácil obtener datos de calidad y en la que exista una mayor probabilidad de encontrar un efecto o una asociación, si existe. Esta actitud tiene algunos inconvenientes, ya que si los criterios son muy estrictos, puede no encontrarse un número suficiente de sujetos o no detectarse un efecto en un determinado subgrupo por no haberse considerado antes. Además, se trata de una población altamente seleccionada, lo que dificulta la generalización o extrapolación de los resultados. Precisamente ésta es una de las principales limitaciones de los ECA, tal como se realizan habitualmente, como base para la elaboración de recomendaciones terapéuticas.

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

No hay un acuerdo sobre cuál de ambas actitudes es la más aconsejable. La elección de una u otra dependerá de la propia actitud del investigador y del objetivo concreto que se desee alcanzar.

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

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

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- Birtwhistle RV, Godwin MS, Delva MD, Casson RI, Lam M, MacDonald SE, et al. Randomised equivalence trial comparing three month and six month follow up of patients with hypertension by family practitioners. *BMJ*. 2004;328:204.
- Djulbegovic B, Lasevic M, Cantor A, Fields KK, Bennett CL, Adams JR, et al. The uncertainty principle and industry-sponsored research. *Lancet*. 2000;356:635-8.
- García Puig J, Mateos Antón F, Gil Aguado A, Barcina Sánchez C. ¿Puede conocerse el «doble ciego» antes de desvelar los códigos de identificación? *Med Clin (Barc)*. 1995;105:437-8.
- Hennekens CH, Eberlein K. A randomized trial of aspirin and beta-carotene among US physicians. *Prev Med*. 1985;14:165-8.
- Lipid Research Clinics Program. The lipid research clinics coronary primary prevention trial results. 1. Reduction in incidence of coronary heart disease. *JAMA*. 1984;251:351-64.
- Marín-Iranzo R, de la Sierra-Iserte A, Roca-Cusachs A, Oliván-Martínez J, Redón-Mas J, Luque-Otero M, et al. Estudio doble ciego de la eficacia y la seguridad de la combinación a dosis fija de 10 mg de enalapril/20 mg de nifrendipino en comparación con el incremento de dosis de amlodipino en pacientes con hipertensión esencial no controlada con 5 mg de amlodipino. *Rev Clin Esp*. 2005;205:418-24.
- Multiple Risk Factor Intervention Trial Research Group. Multiple risk factor intervention trial. *JAMA*. 1982;248:1465-77.
- Pierce M, Lundy S, Palanisamy A, Winning S, King J. Prospective randomised controlled trial of methods of call and recall for cervical cytology screening. *BMJ*. 1989;299:160-2.
- Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med*. 1989;321:129-35.

## BIBLIOGRAFÍA

- Argimon JM. La ausencia de significación estadística en un ensayo clínico no significa equivalencia terapéutica. *Med Clin (Barc)*. 2002;118:701-3.
- Bakke OM, Carné X, García Alonso F. *Ensayos clínicos con medicamentos: fundamentos básicos, metodología y práctica*. Barcelona: Doyma; 1994.
- Boissel JP. Planning of clinical trials. *J Intern Med*. 2004;255:427-38.
- Charlton BG. Understanding randomized controlled trials: explanatory or pragmatic? *Fam Pract*. 1994;11:243-4.
- Chow SC, Shao J. A note on statistical methods for assessing therapeutic equivalence. *Control Clin Trials*. 2002;23:515-20.
- Cobos A. *Ensayos clínicos aleatorios*. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editores. Publicación científica biomédica: cómo escribir y publicar un artículo de investigación. Barcelona: Elsevier; 2010. p. 145-60.
- Dasgupta A, Lawson KA, Wilson JP. Evaluating equivalence and noninferiority trials. *Am J Health Syst Pharm*. 2010;67:1337-43.
- Djulbegovic B, Clarke M. Scientific and ethical issues in equivalence trials. *JAMA*. 2001;285:1206-8.
- Elie C, De Rycke Y, Jais JP, Marion-Gallois R, Landais P. Methodological and statistical aspects of equivalence and non inferiority trials. *Rev Epidemiol Santé Publique*. 2008;56:267-77.
- Ellenberg SS, Temple R. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 2: practical issues and specific cases. *Ann Intern Med*. 2000;133:464-70.
- Fergusson D, Aaron SD, Guyatt G, Hebert. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ*. 2002;325:652-4.
- Fleming TR. Design and interpretation of equivalence trials. *Am Heart J*. 2000;139:S171-6.
- Fletcher RH. Evaluation of interventions. *J Clin Epidemiol*. 2002;55:1183-90.
- García Alonso F, Guallar E, Bakke OM, Carné X. El placebo en ensayos clínicos con medicamentos. *Med Clin (Barc)*. 1997;109:797-801.
- García López FJ, Gutiérrez Bezón S, Galende Domínguez I, Avendaño Solá C. Evaluación de calidad de los ensayos clínicos: justificación, utilidad e inconvenientes. *Med Clin (Barc)*. 1999;112 Suppl 1:35-42.
- Gibaldi M, Sullivan S. Intention-to-treat analysis in randomized trials: who gets counted? *J Clin Pharmacol*. 1997;37:667-72.
- Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol*. 2003;3:28.
- Gomberg-Maitland M, Frison L, Halperin JL. Active-control clinical trials to establish

- equivalence or noninferiority: methodological and statistical concepts linked to quality. *Am Heart J.* 2003;146:398-403.
- Green SB. Design of randomized trials. *Epidemiol Rev.* 2002;24:4-11.
- Guyatt G, Rennie D, editors. *Users' Guides to the Medical Literature: a manual for evidence-based clinical practice.* Chicago: American Medical Association; 2002.
- Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ.* 1999;319:670-4.
- Jiménez Villa J, Argimon Pallàs JM. Ensayos de equivalencia y no-inferioridad. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. *Publicación científica biomédica: cómo escribir y publicar un artículo de investigación.* Barcelona: Elsevier; 2010. p. 161-74.
- Jiménez Villa J, Argimon Pallàs JM. Análisis e interpretación de un ensayo clínico. *Formación Médica Continuada FMC.* 2009;16:261-71.
- Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ.* 1996;313:36-9.
- Kienle G, Kiene H. The powerful placebo effect: fact or fiction? *J Clin Epidemiol.* 1997;50:1311-8.
- Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trials.* 2000;21:167-89.
- Manchikanti L, Hirsch JA, Smith HS. Evidence-Based Medicine,
- Systematic Reviews, and Guidelines in Interventional Pain Management: Part 2: Randomized Controlled Trials. *Pain Physician.* 2008;11:717-73.
- McMahon AD. Study control, violators, inclusion criteria and defining explanatory and pragmatic trials. *Stat Med.* 2002;21:1365-76.
- 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.
- 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.
- Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomised trials. The CONSORT Group. *Lancet.* 2001;357:1191-4.
- Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ.* 2001;165:1339-41.
- Pablos Méndez A, Barr G, Shea S. Run-in periods in randomized trials: implications for the analysis of results in clinical practice. *JAMA.* 1998;279:222-5.
- Peduzzi P, Henderson W, Hartigan P, Lavori P. Analysis of randomized controlled trials. *Epidemiol Rev.* 2002;24:26-38.
- Pocock SJ. When to stop a clinical trial. *BMJ.* 1992;305:235-40.
- Roland M, Torgerson BJ. What are pragmatic trials? *BMJ.* 1998;316:285.
- Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine: how to practice and teach EBM.* 2nd ed. London: Churchill Livingstone; 2000.
- Sacristán JA, Soto JA, Galende I. *Evaluación crítica de ensayos clínicos.* Med Clin (Barc). 1993;100:780-7.
- Schulz KF, Altman DG, Moher D. Allocation concealment in clinical trials. *JAMA.* 2002;288:2406-7.
- Schultz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA.* 1995;273:408-12.
- Scott IA. Non-inferiority trials: determining whether alternative treatments are good enough. *Med J Aust.* 2009;190:326-30.
- Stanley K. Design of randomized controlled trials. *Circulation.* 2007;115:1164-9.
- Stanley K. Evaluation of randomized controlled trials. *Circulation.* 2007;115:1819-22.
- Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. *Ann Intern Med.* 2000;133:455-63.
- Walach H, Sadaghiani C, Dehm C, Bierman D. The therapeutic effect of clinical trials: understanding placebo response rates in clinical trials: a secondary analysis. *BMC Med Res Methodol.* 2005;5:26.

# Capítulo

# 6

## Estudios experimentales II: otros diseños

En el capítulo anterior se han presentado las principales características del diseño experimental más utilizado en investigación clínica (el ensayo clínico aleatorio paralelo), en su finalidad más frecuente (la comparación de la eficacia de dos intervenciones terapéuticas). Sin embargo, su utilización para evaluar la eficacia de otro tipo de intervenciones, como una medida preventiva o una prueba diagnóstica, requiere realizar algunas consideraciones específicas. Además, existen otros diseños de ensayo clínico aleatorio (ECA) diferentes del paralelo, como el diseño cruzado, el factorial o el secuencial, que tienen ventajas importantes en determinadas situaciones. También existen otros estudios experimentales cuyas principales características deben conocerse. Estos aspectos son los que se abordan en este capítulo.

### DESARROLLO CLÍNICO DE MEDICAMENTOS

El desarrollo de un fármaco es una actividad compleja que requiere gran cantidad de recursos y tiempo. Antes de comercializar un nuevo fármaco, se debe realizar un extenso proceso de experimentación que ha de presentarse para su aprobación a las agencias reguladoras de los países en los que se desea comercializar. Aproximadamente, de cada 5.000 compuestos que entran en fase preclínica, tan sólo cinco pasan a la fase clínica y solamente uno llega a obtener la autorización para su comercialización.

La primera etapa del proceso es la investigación preclínica, durante la cual se realizan estudios de laboratorio y en animales para demostrar la actividad biológica del compuesto en una determinada enfermedad y evaluar su seguridad. Estos estudios

duran aproximadamente 3 años y medio. Sólo si el fármaco se muestra seguro para administrarlo a los seres humanos, se pasa a la etapa de desarrollo clínico, que consiste generalmente en cuatro fases, comenzando por estudios a pequeña escala para determinar su seguridad ([tabla 6.1](#)). Posteriormente, se amplía a ensayos más grandes para probar su eficacia. La fase I comienza con la primera administración del nuevo fármaco a seres humanos y suele durar aproximadamente un año. Comprende estudios sin objetivo terapéutico que suelen llevarse a cabo en voluntarios sanos normales con la finalidad de estudiar el perfil de seguridad del fármaco, los rangos de seguridad de la dosificación, los datos farmacológicos, la absorción, la distribución, la metabolización, la excreción y la duración de su acción. Sólo si se demuestra que su administración a seres humanos es razonablemente segura (algunos efectos adversos pueden ser aceptables, dependiendo de la gravedad de la enfermedad), puede seguirse adelante. En esta fase también se realiza la denominada *prueba de concepto*, que permite determinar si existen indicios razonables de que el fármaco puede ser eficaz en la indicación seleccionada.

En la fase II se empieza a explorar la eficacia terapéutica en pacientes. Habitualmente de diseñan estudios de aproximadamente 100 a 300 pacientes voluntarios para confirmar la seguridad, determinar la eficacia en seres humanos a corto plazo y ayudar a determinar algunos parámetros como la dosificación para los ensayos posteriores. Típicamente se realizan ensayos controlados con placebo y doble ciego. Estos ensayos suelen durar alrededor de 2 años. El uso adecuado de la información obtenida en los ensayos de las fases preclínica y I permite optimizar los diseños de los ensayos de la fase II, de modo que se puede demostrar claramente la eficacia

**Tabla 6.1 Clasificación de los estudios en función del momento del desarrollo de un fármaco**

Fases	Características
<b>Preclínica</b>	<ul style="list-style-type: none"> <li>• <b>Duración:</b> 3,5 años</li> <li>• <b>Objetivo principal:</b> determinar la seguridad y la actividad biológica</li> <li>• <b>Diseño:</b> estudios de laboratorio y animales</li> </ul>
<b>Fase I</b>	<ul style="list-style-type: none"> <li>• Primer estadio de la prueba de un nuevo fármaco en la especie humana</li> <li>• <b>Duración:</b> 1 año</li> <li>• <b>Objetivo principal:</b> evaluar la seguridad y la dosificación del fármaco en humanos. A veces también se obtienen datos farmacocinéticos y farmacodinámicos</li> <li>• <b>Sujetos:</b> voluntarios sanos (de 20 a 80 individuos)</li> <li>• <b>Diseño:</b> habitualmente son estudios no controlados</li> </ul>
<b>Fase II</b>	<ul style="list-style-type: none"> <li>• <b>Duración:</b> 2 años</li> <li>• <b>Objetivo principal:</b> conocer la farmacocinética y la farmacodinamia (búsqueda de dosis, mecanismo de acción farmacológica, relaciones dosis/respuesta) y los efectos adversos</li> <li>• <b>Sujetos:</b> pacientes potenciales (de 100 a 300 individuos), aunque inicialmente también puede llevarse a cabo en voluntarios sanos</li> <li>• <b>Diseño:</b> estudios no controlados y ensayos clínicos aleatorios controlados con placebo, bajo criterios de selección muy estrictos</li> </ul>
<b>Fase III</b>	<ul style="list-style-type: none"> <li>• Última fase de la evaluación de un medicamento antes de su comercialización</li> <li>• <b>Objetivo principal:</b> evaluar la eficacia y relación beneficio/riesgo en comparación con otras alternativas terapéuticas disponibles, o con un placebo si no hay tratamiento disponible. Permite establecer la eficacia del nuevo fármaco e identificar y cuantificar los efectos indeseables más frecuentes</li> <li>• <b>Sujetos:</b> pacientes (de 1.000 a 3.000 individuos)</li> <li>• <b>Diseño:</b> ensayos clínicos aleatorios</li> </ul>
<b>Fase IV</b>	<ul style="list-style-type: none"> <li>• Estudios realizados después de la comercialización del fármaco</li> <li>• <b>Objetivo principal:</b> evaluar mejor el perfil de seguridad, las posibles nuevas indicaciones o nuevas vías de administración, la eficacia en las condiciones habituales de uso (efectividad) o en grupos especiales</li> <li>• <b>Sujetos:</b> pacientes</li> <li>• <b>Diseño:</b> ensayos clínicos aleatorios y estudios observacionales</li> </ul>

(o ineficacia) de un fármaco antes de realizar los costosos ensayos de la fase III.

Los ensayos clínicos de fase III se diseñan para probar la eficacia y la seguridad de un fármaco a largo plazo. Son generalmente estudios a doble ciego, controlados con placebo, e incluyen habitualmente de 1.000 a 3.000 pacientes en un período de cerca de 3 años. Aunque el número de pacientes incluidos puede ser elevado, los riesgos son relativamente mínimos, debido a que las pruebas anteriores establecen su seguridad. El gran alcance de estos ensayos da a los investigadores la oportunidad de demostrar la eficacia y la seguridad del medicamento, así como identificar algunos efectos secundarios raros del tratamiento, si los hubiera.

Toda la información obtenida se recopila en un informe o expediente de registro, que se presenta a las autoridades sanitarias para solicitar la aprobación de la comercialización del fármaco.

Tras la comercialización, se debe continuar remitiendo informes periódicos a las autoridades

reguladoras, incluyendo todos los casos de reacciones adversas y los registros de control de calidad apropiados. También pueden realizarse estudios posteriores a la comercialización del medicamento (fase IV), que suelen ser de gran tamaño, a menudo con diseños observacionales, dirigidos a conocer mejor su perfil de seguridad, sus efectos a largo plazo y otros aspectos relacionados con su utilización.

## EVALUACIÓN DE TRATAMIENTOS NO FARMACOLÓGICOS

Aproximadamente el 25% de los ensayos clínicos que se publican evalúan intervenciones no farmacológicas, como la cirugía, la rehabilitación, la psicoterapia, las terapias conductuales o las terapias alternativas. Dado que suele tratarse de intervenciones complejas cuyos resultados dependen en gran medida de la

experiencia previa de los profesionales y centros que las realizan, la evaluación de su eficacia entraña dificultades metodológicas específicas que, si no se tienen en cuenta al diseñar y ejecutar el estudio, pueden tener un impacto importante sobre la estimación de sus efectos.

Existen evidencias que sugieren que, cuanto mayor es la experiencia de un centro o de un profesional, mejores resultados se obtienen, de manera que, si no es similar en los grupos que se comparan, la estimación del efecto puede estar sesgada. Por ello, deben establecerse y justificarse los criterios de selección, no solamente de los pacientes, sino también de los centros y profesionales participantes, teniendo en cuenta su experiencia y el volumen de intervenciones de este tipo que realizan habitualmente, así como realizar un procedimiento de asignación aleatoria que garantice la comparabilidad de los grupos.

Por otro lado, los tratamientos no farmacológicos suelen ser intervenciones complejas con múltiples componentes que pueden influir sobre el efecto del tratamiento. Por este motivo, no debe considerarse únicamente la intervención en su conjunto, sino también cada uno de sus componentes en particular (por ejemplo, en el caso de una intervención quirúrgica, los cuidados preoperatorios, los que se realizan durante el acto quirúrgico y los que reciben los pacientes en el postoperatorio). Estos aspectos también son importantes en relación con la intervención de comparación, en especial cuando se trata de cuidados habituales, dado que pueden variar mucho entre centros y profesionales. El uso de placebo como comparador en estos estudios puede infravalorar el efecto del tratamiento.

**Ejemplo 6.1.** Suponga que se desea diseñar un ensayo para evaluar la eficacia de la acupuntura. Además del uso de agujas, esta intervención tiene distintos componentes, como el proceso de hablar con el paciente y escucharlo. Dentro de las sesiones de tratamiento, estos factores característicos son distintos, pero no distinguibles de otros elementos secundarios, como la empatía. Estos hechos tienen consecuencias importantes para el diseño de ensayos clínicos. Si se utiliza un placebo, una característica definitoria de los ensayos clínicos con una finalidad explicativa, se está asumiendo que el empleo de agujas es el único componente del tratamiento. Por lo tanto, los participantes en el grupo de control reciben todo excepto el uso de agujas. Sin embargo, si el tratamiento tiene otros componentes característicos, el uso de placebo puede subestimar el efecto terapéutico total. Un ensayo clínico con placebo sólo es apropiado para comparar dos técnicas de acupuntura, pero no para valorar el efecto total característico de la acupuntura (Paterson y Dieppe, 2005).

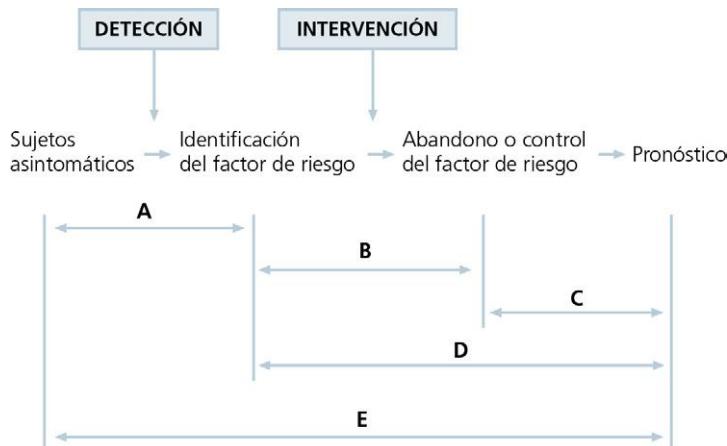
Al existir múltiples componentes y gran variabilidad entre los centros y los profesionales que prestan la atención, debe tratarse de estandarizar la intervención lo máximo posible, especialmente en los ensayos diseñados con una perspectiva explicativa. También deben preverse técnicas para mejorar la metodología o la técnica de intervención por parte de los profesionales. La posible variabilidad existente debe tenerse en cuenta al realizar la asignación aleatoria para garantizar en lo posible la comparabilidad de los grupos, lo cual es especialmente importante si se tiene en cuenta que en la mayoría de los casos no podrán utilizarse técnicas de enmascaramiento.

## EVALUACIÓN DE LA EFICACIA DE MEDIDAS PREVENTIVAS

Los ensayos clínicos que evalúan la eficacia de una medida preventiva presentan algunas diferencias respecto de los que estudian la eficacia de un tratamiento. La primera es que habitualmente se estudian individuos presuntamente sanos. Ello implica que el número de personas que desarrollarán la enfermedad es muy bajo, y quizás tras un largo período de tiempo, mientras que las complicaciones de una enfermedad se pueden detectar en una alta proporción de individuos enfermos en un tiempo relativamente corto. En consecuencia, los ensayos que evalúan una medida preventiva suelen requerir un mayor número de individuos y un seguimiento más largo, lo que comporta importantes problemas organizativos.

En muchas ocasiones, los estudios que evalúan medidas de prevención secundaria, como la detección precoz de una enfermedad, se limitan a valorar las características de la prueba de identificación de los sujetos con el factor de riesgo o la enfermedad (fig. 6.1A), considerando que la detección precoz mejora el pronóstico de estos sujetos. En estos estudios se asume que existe una intervención eficaz, y que esta eficacia es tanto mayor cuanto antes se aplique, lo que no siempre es cierto. Además, en muchas ocasiones se ha demostrado que la intervención es eficaz para controlar el factor de riesgo, pero no que ello redunde en un mejor pronóstico de los sujetos. Por tanto, la evaluación de la eficacia de una medida de detección precoz debería incluir estudios que comprobaran si su aplicación seguida de la intervención mejora el pronóstico, lo que debería realizarse mediante un ECA (fig. 6.1E).

Si la eficacia de esta prueba de detección precoz no se evalúa mediante un ECA, se puede concluir que la nueva prueba mejora el pronóstico incluso cuando el tratamiento se deriva de haber detectado que un caso es ineficaz. La simple observación puede



- A. Evaluación de la utilidad de la prueba de detección precoz para la identificación de los sujetos con el factor de riesgo.
- B. Evaluación de la eficacia de la intervención sobre el abandono o el control del factor de riesgo.
- C. Evaluación de la relación entre el abandono o el control del factor de riesgo y el pronóstico de los sujetos.
- D. Evaluación de la eficacia de la intervención sobre el pronóstico de los sujetos.
- E. Evaluación de la eficacia de la detección precoz y la posterior intervención sobre el pronóstico de los sujetos.

**Figura 6.1** Evaluación de la eficacia de una actividad preventiva.

llevar a engaño en multitud de ocasiones porque, además de las limitaciones inherentes a los estudios observacionales, cuando se evalúan actividades preventivas hay que tener en cuenta la posibilidad de tres sesgos específicos.

En primer lugar, el llamado *sesgo de participación*. Dado que, en general, las personas que aceptan y reciben una medida preventiva suelen gozar de mejor salud que aquellas que la rechazan o no tienen acceso a ella, si se realizara un estudio observacional que comparara los sujetos que han recibido la medida preventiva con los que la han rechazado, podría obtenerse una conclusión errónea favorable a la medida preventiva. De hecho, este beneficio aparente puede deberse a otros muchos factores que no han podido controlarse en un estudio observacional, entre los que uno de los más importantes es la autoselección de los sujetos que forman el grupo de estudio. El mejor modo de tener una cierta garantía de que no se produce este sesgo de participación es la asignación al azar de los individuos a los grupos de estudio.

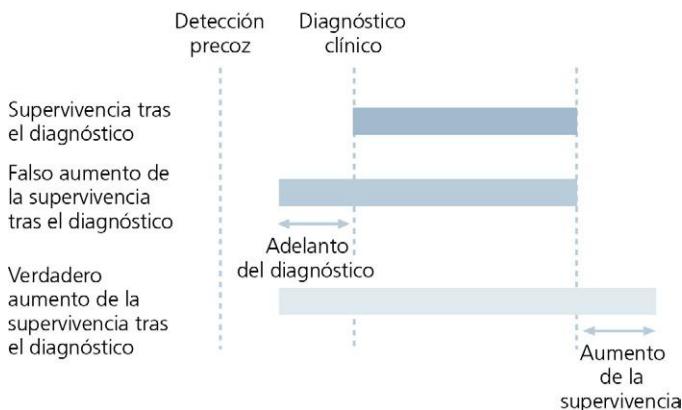
En segundo lugar, el *sesgo por adelanto en el diagnóstico* (*lead time bias*). El período de latencia de una enfermedad se define como el tiempo transcurrido entre su inicio biológico y la aparición de los signos o los síntomas que permitirían su diagnóstico. Durante este período, el problema de salud puede ser detectado a través de la realización de una actividad preventiva antes del

momento en que lo habría sido si se hubiera dejado a su evolución natural. Cuando se evalúa la eficacia de una prueba de detección precoz, debe tenerse en cuenta este tiempo de adelanto del diagnóstico (*lead time*) para evitar sobreestimar los beneficios obtenidos (fig. 6.2).

**Ejemplo 6.2.** Supongamos un estudio hipotético en el que la supervivencia media del grupo de sujetos sometidos a cribado es de 7 años, mientras que la del grupo control es de 5 años. En apariencia ha existido un aumento de 2 años de la supervivencia media. Sin embargo, debido a la existencia de un intervalo de detección, lo que puede haber ocurrido es que se haya adelantado en 2 años el diagnóstico de la enfermedad, sin que en realidad se haya mejorado el pronóstico de los sujetos. De hecho, lo que se ha conseguido es aumentar el tiempo de enfermedad.

El tercer problema potencial es el *sesgo de duración de la enfermedad* (*length time bias*). Puede producirse porque el cribado tiene mayor probabilidad de detectar los casos de progresión más lenta, que quizás sean menos graves y tengan mejor pronóstico, ya que su período presintomático es más largo.

La realización de un ECA, dado que lleva a cabo una asignación aleatoria de los sujetos para conseguir grupos de características similares, previene la aparición de estos sesgos.



**Figura 6.2** Sesgo por adelanto del diagnóstico (*lead time bias*).

## EVALUACIÓN DE LA EFICACIA DE PRUEBAS DIAGNÓSTICAS

La evaluación de la eficacia de una prueba diagnóstica puede abordarse bajo diferentes enfoques. Por un lado, puede diseñarse un estudio con la finalidad de determinar la capacidad de la prueba para distinguir entre los sujetos que padecen un problema de salud y los que no lo padecen. Este diseño corresponde a un estudio descriptivo de las características de la prueba, comparándola con un estándar de referencia, y se aborda en el capítulo correspondiente.

El segundo enfoque consiste en evaluar los beneficios y riesgos asociados al uso de la prueba. Para ello, debe considerarse la prueba diagnóstica como una intervención sanitaria, y diseñar un ECA en que los sujetos sean asignados aleatoriamente a dos grupos, a uno de los cuales se aplica una estrategia diagnóstica que incluye la prueba en estudio, mientras que al otro grupo se le aplica una estrategia alternativa (grupo de comparación). Se realiza un seguimiento de todos los sujetos para determinar la frecuencia de aparición de los resultados de interés clínico en ambos grupos y compararla, tal como se ha comentado en el caso de la evaluación de la eficacia de una actividad preventiva.

## ASIGNACIÓN POR GRUPOS

Habitualmente, en un ECA cada sujeto se asigna de forma individual a uno de los grupos que se comparan, recibe directamente la intervención asignada y sobre él se mide la respuesta observada. Sin embargo, en ocasiones, esta asignación individual no es posible o no resulta adecuada, por lo que se recurre a

una asignación por grupos (*clusters*), como ocurre, por ejemplo, cuando se asignan zonas geográficas para desarrollar un programa de cribado de cáncer de mama, o escuelas para recibir un determinado programa educativo de promoción de la salud, o cuando se evalúan los efectos de una guía de práctica clínica o una intervención de educación médica, de manera que la unidad de asignación la constituyen los profesionales y centros sanitarios, y no los pacientes.

Este diseño también es útil cuando se pretende evitar la posible *contaminación* que puede producirse entre las intervenciones al realizarse en un mismo emplazamiento. Por ejemplo, si se desea comparar el efecto del consejo individualizado para la modificación de los hábitos de vida de los pacientes, formando y entrenando a los profesionales sanitarios para proporcionarlo. Si se asignara a los pacientes de forma individual, un mismo profesional tendría pacientes de ambos grupos, pero, dado que ha sido formado y entrenado para dar consejo, difícilmente podría proporcionar la atención habitual (como si no hubiera recibido la formación específica) a los sujetos del grupo control. Este fenómeno podría evitarse realizando una asignación por profesionales (o por centros de salud).

**Ejemplo 6.3.** Eccles et al (2001) eligieron aleatoriamente a 244 equipos de atención primaria con el fin de evaluar dos estrategias de formación para reducir el número de solicitudes de exploraciones radiológicas en los pacientes que consultaban por problemas de rodilla o en la espalda. La variable de respuesta que utilizaron fue el número de peticiones de radiología por 1.000 pacientes y año.

Estos estudios suelen ser más difíciles de diseñar y ejecutar que los ECA habituales, y deben tenerse en

cuenta algunos aspectos que los hacen más vulnerables en relación con la introducción de sesgos. La asignación de los grupos debe realizarse de forma aleatoria y con el mismo rigor que si fuera individual. También puede introducirse un sesgo de selección si los profesionales de cada uno de los grupos identifican y reclutan a los pacientes de forma diferente. Puffer et al (2003) encontraron problemas en este último aspecto que podrían haber introducido un sesgo en 14 de los 36 ensayos con asignación por grupos publicados en *British Medical Journal (BMJ)*, *Lancet* y *New England Journal of Medicine* entre enero de 1997 y octubre de 2002.

El principal problema de estos estudios es que no puede asumirse que el resultado de cada paciente es independiente del de cualquier otro (p. ej., el manejo de dos sujetos por un mismo profesional es probable que sea más similar que si fueran atendidos por dos profesionales diferentes), por lo que su potencia estadística es menor que la de un ECA habitual con un número de pacientes parecido. Este hecho debe tenerse en cuenta al determinar el tamaño necesario de la muestra, que puede llegar a ser muy superior al de un ECA habitual si existe una gran variabilidad entre las unidades de asignación y poca variabilidad en el interior de cada una de ellas.

Por otro lado, este mismo hecho tiene repercusiones importantes en el análisis de los resultados, que deberá realizarse de forma agregada, ya que la unidad de análisis debe corresponder a la unidad de asignación (zona geográfica, escuela, centro de salud, profesional, etc.), ya que, si se ignora el hecho de que la asignación se ha realizado por grupos, y se analizan los pacientes como un ECA habitual, se produce una disminución de los valores de significación estadística y un estrechamiento de los intervalos de confianza, aumentando la probabilidad de detectar diferencias significativas y obtener conclusiones erróneas. Una forma sencilla de analizar estos estudios es construir un estadístico sintético para cada una de las unidades de asignación, y analizarlos, preferiblemente realizando alguna ponderación según su tamaño. Sin embargo, las técnicas de análisis habituales suelen ser ineficientes y es preferible recurrir a técnicas más específicas.

**Ejemplo 6.4.** En el Reino Unido se realizó un estudio en 34 consultas de atención primaria que derivaban a los pacientes para exploraciones radiológicas a un mismo centro. Los pacientes fueron asignados aleatoriamente a dos grupos (Oakeshott et al, 1994). Las consultas de uno de los grupos recibieron un documento con las guías del Royal College of Radiologists adaptadas a la atención primaria con los criterios recomendados para la derivación. No se

realizó ninguna intervención sobre las consultas del grupo control. Dado que todos los pacientes de una misma consulta recibieron la misma intervención, la unidad experimental fue la consulta. La medida de la respuesta fue el porcentaje de exploraciones radiológicas solicitadas que podían considerarse adecuadas de acuerdo con las guías.

## ENSAYO CLÍNICO SECUENCIAL

En el diseño habitual de un ECA se determina el número de sujetos necesario para tener la suficiente potencia estadística que permita detectar una diferencia determinada con un nivel de significación fijado, asumiendo que el análisis se realizará una vez que se hayan recogido los datos de todos los sujetos.

En ocasiones, por motivos éticos y económicos, puede ser importante realizar alguna monitorización de los datos del estudio, con la intención de finalizarlo tan pronto como exista la suficiente evidencia de que uno de los tratamientos es superior al otro, o bien de que ambas alternativas que se comparan son iguales. Sin embargo, la repetición de análisis estadísticos a medida que se van acumulando datos tiene ciertos problemas. Por un lado, la probabilidad de cometer un error tipo I, es decir, de concluir erróneamente que el tratamiento de estudio es diferente del de comparación, aumenta con el número de análisis intermedios que se realicen (p. ej., si se realizan cuatro análisis intermedios, la probabilidad de cometer un error tipo I se sitúa alrededor del 14%). Por otro lado, el análisis final del estudio también debe realizarse teniendo en cuenta la realización previa de los análisis intermedios.

En estas situaciones, se utilizan diseños específicos que se engloban bajo la denominación de *métodos secuenciales*. Se trata de ensayos en los que el tamaño de la muestra no está predeterminado, sino que depende de las observaciones que se realizan. Su característica principal es la definición de una regla de finalización explícita en la que se establece la forma en que la decisión de finalizar el estudio depende de los resultados obtenidos hasta ese momento. El cuadro 6.1 muestra los cuatro elementos clave de este tipo de estudios, de los que los dos primeros son comunes con los ECA tradicionales.

El diseño de estos estudios requiere la monitorización de un criterio estadístico que sintetice la diferencia entre los tratamientos a lo largo del estudio. Si el valor de este estadístico excede un valor crítico prefijado, el ensayo se finaliza y se rechaza la hipótesis nula de ausencia de diferencia. Si el valor del estadístico no sobrepasa dicho valor crítico, no

**Cuadro 6.1 Elementos clave para la realización de un diseño secuencial**

1. Parámetro que exprese la ventaja del tratamiento experimental sobre el control. Se trata de una característica desconocida de la población sobre la que pueden realizarse hipótesis y de la que pueden obtenerse estimaciones
2. Criterio estadístico que exprese la ventaja del tratamiento experimental sobre el control, a partir de la muestra de datos disponibles en un análisis intermedio, y un segundo criterio estadístico que exprese la cantidad de información sobre la diferencia entre los tratamientos contenida en dicha muestra
3. Regla de finalización (*stopping rule*) que determine si el análisis intermedio debe ser el último, y, en caso afirmativo, si puede concluirse que el tratamiento experimental es mejor o peor que el control, o bien si no se ha establecido la existencia de una diferencia entre los tratamientos
4. Método de análisis válido para el diseño específico utilizado, que proporcione un valor de *p* y una estimación puntual y por intervalo

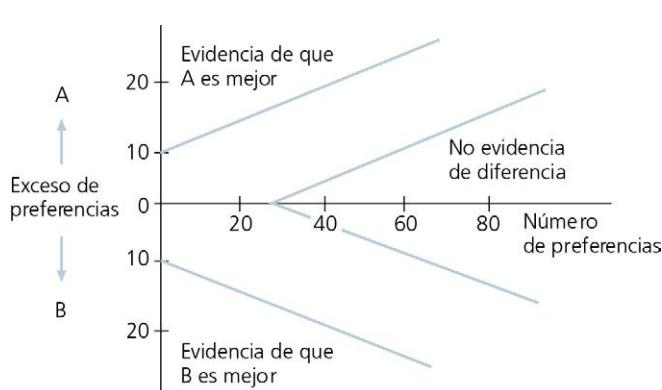
Modificada de Whitehead, 1999.

existe todavía la suficiente evidencia para alcanzar una conclusión y el estudio continúa. Sin embargo, la decisión de finalizar el estudio debe tomar en consideración otros aspectos adicionales, como los efectos secundarios, la facilidad de administración, el coste, las evidencias procedentes de otras fuentes, etc., de manera que el criterio estadístico no sea el único determinante de dicha decisión.

Los análisis intermedios pueden realizarse después de conocer el resultado de cada uno de los pacientes incluidos o bien de grupos de sujetos. Lógicamente, si solamente se realizan uno o dos análisis intermedios, se reduce la posibilidad de detectar precozmente una diferencia y se retrasa la posible finalización del estudio. Sin embargo, dado que la realización de estos análisis complica el estudio, en la práctica se recomienda realizar entre cuatro y ocho análisis intermedios. Una vez finalizado el estudio, debe efectuarse un análisis final que tenga en cuenta el carácter secuencial del ensayo, ya que la utilización de los métodos tradicionales conduce a valores de *p* demasiado pequeños, estimaciones puntuales excesivamente grandes e intervalos de confianza muy estrechos.

Así, por ejemplo, en los *diseños secuenciales de preferencias por parejas*, los pacientes se incluyen de dos en dos, de modo que cada uno de ellos recibe uno de los tratamientos de forma aleatoria, y se determina cuál de los dos responde mejor. A medida que progresa el estudio, se van acumulando preferencias a favor de un tratamiento u otro, que se van representando en un gráfico (fig. 6.3). El ensayo finaliza tan pronto como los resultados alcanzan uno de los límites prefijados por la regla de finalización: si se alcanza el límite superior, se concluye que A es mejor; si se alcanza el límite inferior, se concluye que B es mejor; y si se alcanza el límite central, se concluye que no se tiene evidencia de que existan diferencias entre los tratamientos.

**Ejemplo 6.5.** Allard et al (1999) utilizaron un diseño secuencial para evaluar la eficacia de las dosis suplementarias de opioides para la disnea de los pacientes oncológicos terminales. Incluyeron 15 parejas de pacientes. En cada pareja, el orden de las intervenciones (administración del 25 o del 50% de su dosis de opioides) fue asignado aleatoriamente. Se midieron



**Figura 6.3** Ensayo secuencial de preferencias por parejas.

la intensidad de la disnea y la frecuencia respiratoria durante un período de 4 horas, y se asignó para cada pareja una preferencia al tratamiento más efectivo. Encotraron un número de preferencias similar (8 y 7), por lo que concluyeron que el 25% de la dosis equivalente de 4 horas puede ser suficiente para reducir la intensidad de la disnea y la taquipnea en pacientes oncológicos terminales con disnea persistente.

Para poder utilizar un diseño secuencial, la respuesta debe poder observarse en un tiempo relativamente corto que permita tomar la decisión de finalizar o no el estudio de forma rápida. Los hipnóticos y analgésicos son ejemplos de fármacos que han sido evaluados, en ocasiones, mediante este diseño.

## ENSAYO CLÍNICO FACTORIAL

El diseño factorial permite al investigador evaluar dos intervenciones o más en un único estudio. En su forma más sencilla para evaluar dos tratamientos (A y B), cada sujeto se asigna aleatoriamente a uno de cuatro grupos (**tabla 6.2**): un grupo recibe ambos tratamientos, dos grupos reciben solamente uno de los tratamientos cada uno, y el cuarto grupo no recibe ningún tratamiento. Si existe enmascaramiento, cada uno de los tres últimos grupos debe recibir un placebo del tratamiento o de los tratamientos que no se administran.

Este diseño es muy útil para evaluar a la vez varios tratamientos con mecanismos de acción y efectos independientes (A tiene la misma eficacia tanto en el grupo que recibe también B como en el que no lo recibe, y viceversa) con el mismo número de individuos que hubiera sido necesario para evaluar un solo tratamiento. En este caso, el análisis consiste en comparar todos los sujetos que han recibido cada uno de los tratamientos con los que no lo han recibido.

**Ejemplo 6.6.** En el estudio americano sobre el efecto de la aspirina en la prevención primaria del infarto de miocardio, los investigadores aprovecharon la misma muestra de individuos para estudiar el efecto del betacaroteno sobre el cáncer (Hennekens y Eberlin, 1985). Los sujetos fueron distribuidos aleatoriamente en cuatro grupos: los asignados al primero de ellos recibieron aspirina más betacaroteno; los del segundo, aspirina más un placebo de betacaroteno; los del tercero, betacaroteno más un placebo de aspirina, y los del cuarto grupo, un placebo de betacaroteno más un placebo de aspirina. Los resultados de todos aquellos que recibieron aspirina se compararon con los de aquellos que recibieron placebo de aspirina, con independencia de que recibieran betacaroteno o placebo. Por otro lado, se compararon los individuos que tomaron betacaroteno con los que tomaron su correspondiente placebo con independencia de que recibieran aspirina o placebo.

En la situación planteada en el ejemplo anterior, el diseño factorial es muy eficiente, ya que es capaz de dar respuesta a dos o más hipótesis en un solo estudio. Permite explorar una hipótesis poco madura (como es el caso del efecto del betacaroteno en la prevención del cáncer), junto con otra que posee una mayor evidencia empírica. Sin embargo, es importante que la evaluación de la segunda hipótesis no altere los criterios de inclusión y exclusión de la principal, ya que, de otro modo, se perderá eficiencia y se complicará mucho la ejecución del estudio.

No es conveniente evaluar más de una hipótesis de forma simultánea cuando se sospecha que una de las intervenciones tiene muchos efectos secundarios o es muy mal tolerada, ya que ello condicionaría un mal cumplimiento de todas las intervenciones. Por otro lado, es imprescindible que ambas intervenciones no interaccionen entre sí.

**Tabla 6.2** Diseño factorial que evalúa dos tratamientos

		Tratamiento A		Total
		Sí	No	
Tratamiento B	Sí	AB	B0	B
	No	A0	00	No B
Total		A	No A	
Análisis de los marginales (ausencia de interacción):		Existencia de interacción:		
<ul style="list-style-type: none"> <li>Eficacia de A: A (celdas AB y A0) frente a No A (celdas B0 y 00).</li> <li>Eficacia de B: B (celdas AB y B0) frente a No B (celdas A0 y 00).</li> </ul>		<ul style="list-style-type: none"> <li>Eficacia de A: celda A0 frente a celda 00.</li> <li>Eficacia de B: celda B0 frente a celda 00.</li> </ul>		

Cuando existe interacción entre los tratamientos, tanto si es sinérgica (A es más eficaz si se administra conjuntamente con B, o viceversa) como antagónica (A es menos eficaz cuando se administra conjuntamente con B, o viceversa), la estrategia de análisis es diferente, y obliga a comparar los resultados de cada una de las celdas. Por ejemplo, la eficacia de A se determina comparando el grupo que ha recibido solamente A con el que no ha recibido ningún tratamiento. De forma similar se calcularía la eficacia de B. La evaluación de la interacción supone comparar los efectos de cada uno de los tratamientos cuando se administra solo o conjuntamente con el otro. De ello se deduce que, cuando se utiliza este diseño para evaluar interacciones, el tamaño muestral debe incrementarse de forma importante.

## ENSAYO CLÍNICO CRUZADO

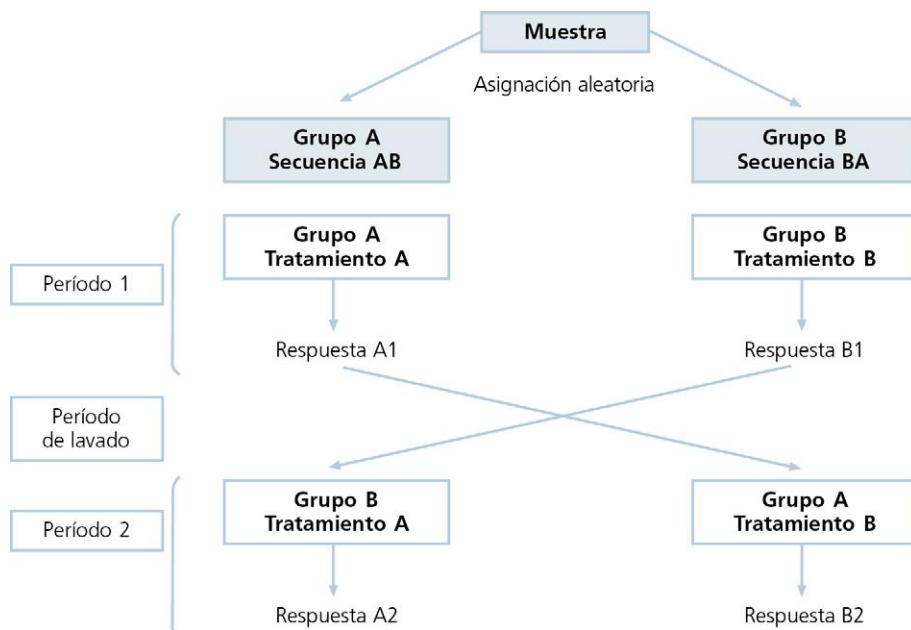
En un ECA paralelo, cada paciente se asigna a un grupo y recibe sólo una de las intervenciones que se comparan. Sin embargo, parece lógico pensar que, si se compararan dos tratamientos, sería preferible administrar ambos a los mismos sujetos, de manera

que cada paciente actuara como su propio control. Dado que la variabilidad intrasujetos es menor que la entre sujetos, la comparación sería más potente y la estimación de la diferencia más precisa. Este tipo de estudios se denominan *ensayos cruzados (crossover)*.

En el caso más sencillo de comparación de dos tratamientos, cada individuo es asignado aleatoriamente a un grupo, que recibe, en un primer período, una de las dos intervenciones y, en un segundo período, la otra (fig. 6.4). Ambos períodos suelen estar separados por una fase de lavado o blanqueo para permitir que el paciente vuelva a su estadio inicial. Este período intermedio debe ser lo suficientemente largo para asegurar que el efecto del tratamiento administrado en el primer período ha desaparecido.

Las características diferenciales entre un estudio en paralelo y uno cruzado se recogen en la tabla 6.3.

**Ejemplo 6.7.** Un estudio tenía por objetivo evaluar la respuesta espirométrica al salbutamol inhalado desde un cartucho presurizado comparada con la de la inhalación con ayuda de una cámara de aerosol, en pacientes con limitación al flujo aéreo (Mayos et al, 1987). El estudio se realizó en dos días sucesivos, siempre por la mañana y a la misma hora, mediante



Efecto tratamiento (columnas):  $(A_1 + A_2)$  frente a  $(B_1 + B_2)$ .  
 Efecto período (filas):  $(A_1 + B_1)$  frente a  $(A_2 + B_2)$ .  
 Efecto secuencia (interacción):  $(A_1 + B_2)$  frente a  $(B_1 + A_2)$ .

**Tabla 6.3 Principales características diferenciales entre un estudio paralelo y uno cruzado**

Estudio paralelo	Estudio cruzado
• Los grupos de estudio y de comparación están formados por sujetos diferentes	• Cada sujeto actúa como su propio control
• Los pacientes se asignan para recibir uno de los tratamientos	• Los sujetos se asignan a la secuencia en que recibirán ambos tratamientos
• Período de blanqueo no necesario	• Período de blanqueo imprescindible

un diseño de doble ciego cruzado. El tratamiento broncodilatador se suspendió 12 horas antes de cada prueba. Cada uno de los días se administró al paciente dos inhalaciones consecutivas de 0,1 mg de salbutamol y dos inhalaciones de placebo, a través de dos cartuchos en apariencia idénticos, uno de forma directa y otro con la cámara de aerosol intercalada:

Día A: cartucho salbutamol y (cartucho placebo + cámara de aerosol).

Día B: cartucho placebo y (cartucho salbutamol + cámara de aerosol).

El orden de distribución del cartucho-cámara de aerosol y del salbutamol-placebo fue aleatorio.

Sin embargo, en la mayoría de las ocasiones este diseño no es factible, ya que habitualmente los tratamientos producen cambios irreversibles en el estado de salud de los pacientes, o a veces deben administrarse durante largos períodos. El diseño cruzado es adecuado cuando se evalúan intervenciones que producen rápidas mejorías en los síntomas de enfermedades crónicas, relativamente estables, y cuyos resultados también desaparecen de forma rápida. No es adecuado en el caso de que la secuencia en que se administran las intervenciones pueda alterar el resultado, o cuando no es posible realizar un período de blanqueo que asegure la total desaparición del efecto de la intervención recibida en primer lugar.

La estructura básica de un diseño cruzado se presenta en la figura 6.4. Los pacientes asignados al grupo A reciben el tratamiento A en el período 1, y el B en el período 2. Los pacientes del grupo B los reciben en el orden inverso. El efecto del tratamiento se obtiene comparando los resultados obtenidos por la intervención A en ambos períodos ( $A_1 + A_2$ ) con los obtenidos por la intervención B ( $B_1 + B_2$ ). Sin embargo, esta comparación no tiene en cuenta que los tratamientos se han administrado en períodos y secuencias diferentes. Además del efecto propio de la intervención, el análisis de los resultados de un ensayo cruzado debe tomar en consideración la posible existencia de los efectos período y secuencia.

### Efecto período

Dado que cada paciente se observa en dos períodos distintos, es importante determinar si ha existido algún cambio entre el primero y el segundo. La enfermedad o síntoma puede progresar, regresar o fluctuar en su gravedad, por lo que es posible encontrar cambios entre los diferentes períodos, con independencia del tratamiento administrado. Una manera de minimizar este efecto es que cada paciente cambie de tratamiento varias veces, recibiéndolos durante varios intervalos de tiempo, de forma que si existe un efecto período, éste quede contrarrestado. Sin embargo, ello dificulta mucho la ejecución del estudio y también el hecho de que los pacientes acepten participar.

La evaluación de la existencia de un efecto período requiere comprobar si existen diferencias entre los resultados obtenidos en el primer período ( $A_1 + B_1$ ) y en el segundo ( $A_2 + B_2$ ).

### Efecto secuencia

Los pacientes no se asignan a un único tratamiento, sino a una secuencia de intervenciones. Si todos los sujetos recibieran las distintas intervenciones siguiendo una misma secuencia, primero A y después B, se asumiría que los efectos del segundo tratamiento B después de A no se diferenciarían de los obtenidos si B se hubiera administrado en primer lugar.

El único modo de evaluar si la secuencia en que se reciben las intervenciones afecta a los resultados es que unos pacientes reciban la secuencia AB y otros la BA. El mejor método es asignar los pacientes de forma aleatoria a cada una de estas secuencias.

Un efecto secuencia puede darse siempre que un individuo se observa más de una vez en períodos distintos. De modo intuitivo, se puede entender como una interacción entre el propio tratamiento y el efecto período que se presenta en las siguientes situaciones:

- Cuando la respuesta en el segundo período está afectada por el tratamiento recibido durante el primero. Es el llamado *efecto residual*. La forma

de prevenirlo es la aplicación de una fase de blanqueo entre ambos períodos, en la que el paciente no recibe ningún tratamiento o, a lo sumo, un placebo. Sin embargo, si el período de blanqueo no es lo bastante largo como para permitir que el efecto del primer tratamiento se haya extinguido, puede existir cierto efecto residual.

- El efecto residual puede estar presente, aunque el período de blanqueo sea lo bastante largo, en el caso de que persistan sus efectos psicológicos durante el segundo período.
- El efecto residual también aparece cuando los individuos no pueden volver a su estado basal después del primer tratamiento. En algunos casos, retirar un medicamento que parece efectivo y esperar a que el paciente vuelva a los valores iniciales puede dar lugar a problemas éticos. En otros, si la intervención aplicada en el primer período cura la enfermedad, no existe la posibilidad de que el paciente retorne a su línea de partida.
- Si existe un fuerte efecto período, el efecto del tratamiento puede verse afectado simplemente por que la enfermedad o el síntoma hayan mejorado o empeorado de un período a otro (la eficacia del tratamiento puede variar según la gravedad de la enfermedad o el valor de la variable de medida).

En cualquier situación, si es probable que después de la administración de los tratamientos en el primer período los pacientes no puedan volver a su estado basal, debe evitarse utilizar un diseño cruzado.

Si existe un efecto secuencia, no es adecuado comparar los tratamientos utilizando todos los datos del estudio. Los resultados del primer período son los únicos que proporcionan una comparación válida, ya que los sujetos han sido asignados aleatoriamente a ambos tratamientos. En cambio, los resultados del segundo período tienen un valor dudoso, ya que los sujetos, aunque habían sido asignados aleatoriamente al inicio, han tenido experiencias diferentes antes de iniciar este período.

Por ello, se recomienda desglosar el análisis de un ensayo cruzado en dos fases. En la primera se evalúa un posible efecto secuencia. Si éste no es significativo (utilizando un criterio poco restrictivo, como, por ejemplo,  $p=0,10$ ), en una segunda fase se analiza el efecto del tratamiento con todos los datos del estudio. Si por el contrario, el efecto secuencia es significativo, se comparan únicamente los resultados del primer período, aunque ello suponga no utilizar buena parte de la información recogida y desperdiciar las teóricas ventajas del diseño cruzado, incluida la economización en el número de sujetos estudiados. Por este motivo, no debe utilizarse este diseño si no se está razonablemente seguro, a partir de estudios previos y

de consideraciones farmacocinéticas, de que no existe ninguna interacción entre el tratamiento y el período, o de que, si existe, es prácticamente despreciable.

## Ventajas

La principal ventaja del diseño cruzado es su eficiencia, ya que cada individuo actúa como su propio control, observándose el mismo número de respuestas que en un estudio en paralelo con la mitad de individuos. Dado que encontrar el número de sujetos suficientes para completar un estudio es un problema frecuente, ésta es una ventaja nada despreciable.

Al evaluar los dos tratamientos en un mismo paciente, la variabilidad intraindividual es menor, lo que permite utilizar pruebas estadísticas para datos apareados que tienen una mayor potencia y, por tanto, se precisa un número todavía menor de sujetos.

## Desventajas

Una de las principales limitaciones deriva, aunque parezca una paradoja, de su eficiencia. Al requerir menos sujetos, cualquier pérdida durante el seguimiento tendrá mayores repercusiones que en un diseño en paralelo, ya que cada paciente aporta mayor cantidad de información. Esto es muy importante porque, debido a que cada uno de los pacientes tiene que recibir los dos tratamientos o más, las pérdidas suelen ser más frecuentes. Además, como cada paciente tiene que completar dos períodos de seguimiento, la duración del estudio es más larga, lo que también facilita las pérdidas.

El diseño cruzado no deberá utilizarse en enfermedades agudas o que cursen con brotes, ni cuando el orden en que se administren las intervenciones pueda alterar el resultado o cuando no sea posible realizar un período de blanqueo que asegure la total desaparición del efecto de la primera intervención.

La duración del período de blanqueo ha de ser corta. Un fármaco que necesita meses para ser eliminado del organismo no es un buen candidato para ser evaluado en un estudio cruzado. Igualmente, este diseño está contraindicado si el tiempo de blanqueo varía mucho de un individuo a otro.

Las ventajas y limitaciones de los ensayos cruzados se resumen en el cuadro 6.2.

## ENSAYOS COMUNITARIOS

El ensayo comunitario es el diseño apropiado para la evaluación de intervenciones de base comunitaria. Por lo tanto, conceptualmente se diferencia de un ensayo clínico en que la intervención no se lleva a cabo separadamente para cada individuo, sino sobre

### Cuadro 6.2 Ventajas y desventajas de un estudio cruzado

#### Ventajas

- Es más eficiente que un estudio en paralelo, ya que requiere un número menor de participantes
- Cada participante es su propio control, por lo que se pueden utilizar técnicas estadísticas para datos apareados, que son más potentes

#### Desventajas

- Tienen mayor duración que los estudios en paralelo
- No puede utilizarse en enfermedades agudas o cuya evolución cursa a brotes
- No se puede aplicar cuando no es posible asegurar la desaparición del efecto de la primera intervención en todos los participantes

la comunidad en su conjunto (p. ej., la fluoración de las aguas de abastecimiento de una comunidad o las recomendaciones sobre estilos de vida realizadas a través de los medios de comunicación).

**Ejemplo 6.8.** En el Minnesota Heart Health Program se seleccionaron tres parejas de comunidades. Cada pareja incluía una comunidad de estudio y una de control que habían sido emparejadas por tamaño, tipo de comunidad y distancia del área metropolitana. Dentro de cada pareja, la asignación al grupo de estudio o al control no fue aleatoria. La intervención iba dirigida a la prevención y el control de la hipertensión arterial, a la promoción de hábitos alimentarios saludables, a la lucha contra el tabaquismo y a la promoción de la práctica regular de ejercicio físico, y se utilizaron múltiples estrategias: implicación de los líderes de la comunidad y de los profesionales sanitarios, medios de comunicación de masas, educación en las escuelas, etc.

En el ejemplo anterior no se realizó una asignación aleatoria dado el reducido número de comunidades de estudio. De hecho, cuando solamente existen dos comunidades, no es importante si la intervención se asigna aleatoriamente o no, ya que las diferencias en la línea basal serán de la misma magnitud (en todo caso, únicamente la dirección de las diferencias se vería afectada). Idealmente, debería incluirse un número de unidades (comunidades) suficiente para que la asignación aleatoria resultara eficaz y tendiera a la distribución equilibrada de las características basales de ambos grupos.

La principal ventaja de los ensayos comunitarios es la elevada capacidad de generalización de sus resultados (validez externa). El fundamento de estos estudios es

desarrollar métodos de reducción de riesgos que sean aplicables en las condiciones reales en las que vive la población y evaluar su eficacia en un contexto que los hace más generalizables que si se llevaran a cabo en un contexto clínico. Otras ventajas se relacionan con el tipo de intervención que evalúan, dada la buena relación coste-eficiencia que presentan, al permitir el uso de los medios de comunicación de masas, la posibilidad de aumentar la eficacia mediante la difusión de la información y el aumento de la comunicación interpersonal, así como la producción de cambios en las estructuras sociales e institucionales que puedan ayudar al mantenimiento de los cambios de conducta.

Sus limitaciones más importantes se relacionan con la dificultad de realizar una inferencia causal, dado que habitualmente existe un reducido número de unidades de intervención, las diferencias en la línea basal de las comunidades que se comparan (en especial, si la asignación no es aleatoria) y la existencia de tendencias temporales en la evolución de la exposición y del resultado, que pueden interferir con la estimación del efecto de la intervención.

En cualquier ensayo de intervención comunitaria con una duración prevista de varios años hay que estimar la tendencia en el tiempo que es probable que se observe en la variable de respuesta en el grupo control. Si la tendencia secular de la variable de respuesta en el grupo control va en la misma dirección que la del grupo de intervención, y es superior a la esperada, será difícil observar un efecto de magnitud suficiente para que el resultado sea estadísticamente significativo. Por esta razón, hay que tener presente esta información en el diseño, para no llevar a cabo un estudio que puede ser muy costoso, y con pocas probabilidades de encontrar el resultado esperado.

Otra desventaja de estos estudios es que su diseño (en especial la estimación del tamaño de la muestra) y el análisis son muy complicados.

## OTROS DISEÑOS EXPERIMENTALES

Además del ECA, existen otros diseños experimentales que se utilizan en investigación clínica (fig. 6.5).

### Ensayos controlados no aleatorios

Los ensayos controlados no aleatorios corresponden a estudios experimentales con un grupo control concurrente, muy similares a los ECA, con la diferencia de que los sujetos se asignan a los grupos de estudio por un mecanismo no aleatorio; por ejemplo, de forma alternativa, por fecha de nacimiento, por número de historia clínica par o impar, etc.



La asignación de los participantes a los grupos del estudio puede realizarse individualmente o por grupos (*clusters*).

Cuando la intervención que se evalúa se realiza sobre una comunidad, se prefiere la denominación de *ensayo comunitario*.

**Figura 6.5** Estudios experimentales.

Dado que la asignación no es aleatoria, existe un mayor riesgo de que la distribución de los factores pronóstico no sea equilibrada entre ambos grupos, de manera que los grupos formados no sean totalmente comparables. Por este motivo, se considera que estos estudios tienen menor calidad metodológica que los ECA.

### Ensayos no controlados

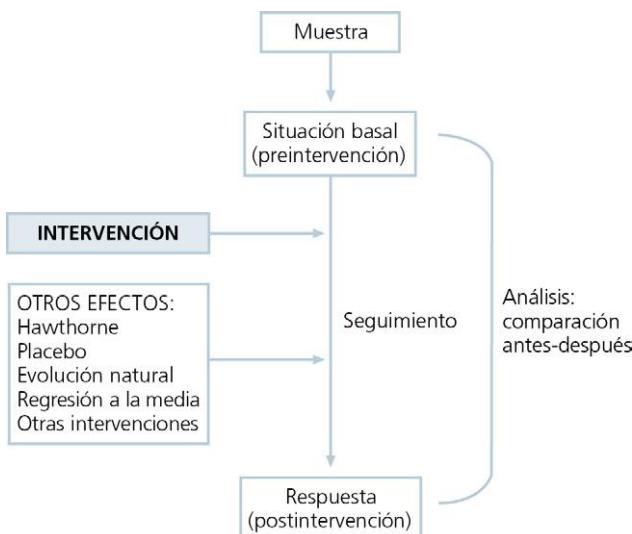
Un estudio de intervención sin grupo control es aquel en el que se administra el fármaco en estudio (o la

intervención de interés) a todos los sujetos que componen la muestra, sin que exista un grupo de comparación concurrente que reciba un placebo u otro tratamiento de referencia. Dado que comparan la respuesta al tratamiento con la situación basal al inicio del estudio, también se denominan *estudios antes-después*.

Al leer un artículo, puede ser difícil diferenciar estos estudios de las series de casos descriptivas, en las que se presentan los resultados observados en un grupo de pacientes que han recibido una intervención que ha sido administrada en el contexto de la práctica clínica habitual, a veces siguiendo un protocolo de tratamiento, pero no en el marco de un proyecto de investigación. Es frecuente que la redacción del artículo no deje suficientemente claro si los investigadores controlaron la intervención de acuerdo con un protocolo de investigación o si se han limitado a describir la experiencia observada en un grupo de sujetos.

El mayor inconveniente de los estudios no controlados radica en la ausencia de un grupo control, por lo que no permiten asegurar que la respuesta observada (cambio en relación con la situación basal) se deba exclusivamente a la intervención de estudio, ya que pueden haber influido otras intervenciones que se hayan producido y otros factores no controlados, como los efectos placebo, Hawthorne o la propia evolución natural de la enfermedad (fig. 6.6). Por este motivo, no se consideran diseños adecuados para evaluar la eficacia de una intervención, sino que su utilidad es fundamentalmente descriptiva.

En la práctica, existen algunas situaciones en las que los ensayos no controlados son los más adecuados por razones éticas y circunstanciales. Por



**Figura 6.6** Esquema de un estudio de intervención no controlado (antes-después).

ejemplo, en un estudio en fase I, en el que en vez de voluntarios sanos se estudian pacientes, puede ser éticamente incorrecto incluir controles tratados con un placebo. Éste es el caso cuando se desarrollan fármacos muy tóxicos, destinados al tratamiento de enfermedades especialmente graves. La seguridad de los enfermos, junto al beneficio potencial, pesan más que las consideraciones metodológicas. Las

fases tempranas del desarrollo de un fármaco tienen por objetivo principal definir las dosis máximas toleradas con pruebas de laboratorio objetivas. Por tanto, en estos casos, la inclusión de un grupo control concurrente no se considera imprescindible. Sin embargo, a medida que progresa el desarrollo de un nuevo fármaco, el empleo de un grupo control concurrente y aleatorizado se hace imprescindible.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- Allard P, Lamontagne C, Bernard P, Tremblay C. How effective are supplementary doses of opioids for dyspnea in terminally ill cancer patients? A randomized continuous sequential clinical trial. *J Pain Symptom Manage.* 1999;17:256-65.
- Birtwhistle RV, Godwin MS, Delva MD, Casson RI, Lam M, MacDonald SE, et al. Randomised equivalence trial comparing three month and six month follow up of patients with hypertension by family practitioners. *BMJ.* 2004;328:204.
- Eccles M, Steen N, Grimshaw J, Thomas L, McNamee P,

- Soutter J, et al. Effect of audit and feedback, and reminder messages on primary care referrals: a randomised trial. *Lancet.* 2001;357:1406-9.
- Hennekens CH, Eberlein K. A randomized trial of aspirin and beta-carotene among US physicians. *Prev Med.* 1985;14:165-8.
- Mayos M, Casan P, Heredia JL, Sanchis J. Cámara de aerosol para pacientes con inhalación incorrecta. *Arch Bronconeumol.* 1987;23:118-22.
- Oakeshott P, Kerry SM, Williams JE. Randomised controlled trial of the effect of the Royal

- College of Radiologists' guidelines on general practitioners' referral for radiographic examination. *Br J Gen Pract.* 1994;44:197-200.
- Paterson C, Dieppe P. Characteristic and incidental (placebo) effects in complex interventions such as acupuncture. *BMJ.* 2005;330:1202-5.
- Puffer S, Torgerson DJ, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ.* 2003;327:785-9.

## BIBLIOGRAFÍA

- Argimon JM. La ausencia de significación estadística en un ensayo clínico no significa equivalencia terapéutica. *Med Clin (Barc).* 2002;118:701-3.
- Argimon JM. Ensayos clínicos con asignación aleatoria de grupos. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica: cómo escribir y publicar un artículo de investigación. Barcelona: Elsevier; 2010. p. 145-60.
- Armitage P, Berry G, Matthews JNS. Statistical Methods in Medical Research. 4th ed. Oxford: Blackwell Science; 2002.
- Atienza AA, King AC. Community-based health

- intervention trials: an overview of methodological issues. *Epidemiol Rev.* 2002;24: 72-9.
- Bland JM, Kerry SM. Trials randomised in clusters. *BMJ.* 1997;315:600.
- Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med.* 2007; 4:61.
- Buring JE. Special issues related to randomized trials of primary prevention. *Epidemiol Rev.* 2002;24:67-71.
- Campbell MK, Elbourne DR, Altman DG, for the CONSORT Group. CONSORT statement:
- extension to cluster randomised trials. *BMJ.* 2004;328:702-8.
- Campbell MK, Fayers PM, Grimshaw JM. Determinants of the intracluster correlation coefficient in cluster randomised trials. *Clin Trials.* 2005;2:99-107.
- Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. *BMJ.* 1998;317:1171-2.
- Campbell MK, Mollison J, Steen N, Grimshaw JM, Eccles M. Analysis of cluster randomized trials in primary care: a practical approach. *Fam Pract.* 2000;17:192-6.
- Cummings P. Carryover bias in crossover trials. *Arch Pediatr Adolesc Med.* 2010;164:703-5.

- Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *Am J Public Health.* 2004;94: 416-22.
- Donner A, Klar N. Statistical considerations in the design and analysis of community intervention trials. *J Clin Epidemiol.* 1996;49:435-9.
- Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: systematic review of recent trials. *BMJ.* 2008;336:876-80.
- Giraudeau B, Ravaud P. Preventing bias in cluster randomised trials. *PLoS Med.* 2009;6:e1000065.
- Hahn S, Puffer S, Torgerson DJ, Watson J. Methodological bias in cluster randomised trials. *BMC Med Res Methodol.* 2005;5:10.
- Jiménez J. Otras extensiones de la declaración CONSORT: intervenciones no farmacológicas, ensayos pragmáticos y resúmenes de ensayos clínicos. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. *Publicación científica biomédica: cómo escribir y publicar un artículo de investigación.* Barcelona: Elsevier; 2010. p. 145-60.
- Kerry SM, Bland JM. Statistics notes: analysis of a trial randomised in clusters. *BMJ.* 1998;316:54.
- Kerry SM, Bland JM. Statistics notes: sample size in cluster randomisation. *BMJ.* 1998;316:549.
- Lancaster GA, Campbell MJ, Eldridge S, Farrin A, Marchant M, Muller S, et al. Trials in primary care: statistical issues in the design, conduct and evaluation of complex interventions. *Stat Methods Med Res.* 2010;19:349-77.
- McAlister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: a systematic review. *JAMA.* 2003;289:2545-53.
- Mills EJ, Chan AW, Wu P, Vail A, Guyatt GH, Altman DG. Design, analysis, and presentation of crossover trials. *Trials.* 2009;10:27.
- Perera R, Heneghan C, Yudkin P. Graphical method for depicting randomised trials of complex interventions. *BMJ.* 2007;334:127-9.
- Tamura RN, Huang X, Boos DD. Estimation of treatment effect for the sequential parallel design. *Stat Med.* 2011;30: 3496-506.
- Todd S, Whitehead A, Stallard N, Whitehead J. Interim analyses and sequential designs in phase III studies. *Br J Clin Pharmacol.* 2001;51:394-9.
- Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *BMJ.* 2001;322: 355-7.
- Vaucher P. Designing phase III or IV trials for vaccines: choosing between individual or cluster randomised trial designs. *Vaccine.* 2009;27:1928-31.
- Wears RL. Statistical methods for analyzing cluster and cluster randomized data. *Acad Emerg Med.* 2002;9:330-41.
- Whitehead J. A unified theory for sequential clinical trials. *Stat Med.* 1999;18:2271-86.

## Estudios de cohortes

El término *cohorte* se utiliza para designar a un grupo de sujetos que tienen una característica o un conjunto de características en común (generalmente la exposición al factor de estudio), y de los que se realiza un seguimiento en el transcurso del tiempo. Puede tratarse de una generación (definida por una misma fecha de nacimiento), un grupo profesional (p. ej., médicos de un país), personas que presentan una determinada exposición (p. ej., mujeres tratadas por cáncer de mama) o característica genética (p. ej., trisomía 21), o una comunidad definida geográficamente (p. ej., los habitantes de la población de Framingham).

Un *estudio de cohortes* es un diseño observacional analítico longitudinal en el que se comparan dos cohortes, o dos grupos dentro de una misma cohorte, que difieren por su exposición al factor de estudio, con el objetivo de evaluar una posible relación causa-efecto. Cuando solamente existe una cohorte que es seguida en el tiempo con la única finalidad de estimar la incidencia con que aparece un determinado problema de salud (desenlace o efecto) o describir su evolución, se trata de un diseño descriptivo longitudinal (ver capítulo 9).

En un estudio de cohortes, los individuos, inicialmente sin la enfermedad o el efecto de interés, forman los grupos en función de su exposición o no al factor o los factores de estudio, y se les realiza un seguimiento durante un período de tiempo comparando la frecuencia con que aparece el efecto o la respuesta en los expuestos y no expuestos (fig. 7.1).

**Ejemplo 7.1.** En el Framingham Heart Study se determinaron los valores de colesterol, presión arterial, hábito tabáquico y otras características en los sujetos de estudio (residentes en la población de Framingham) en un momento en que en apariencia

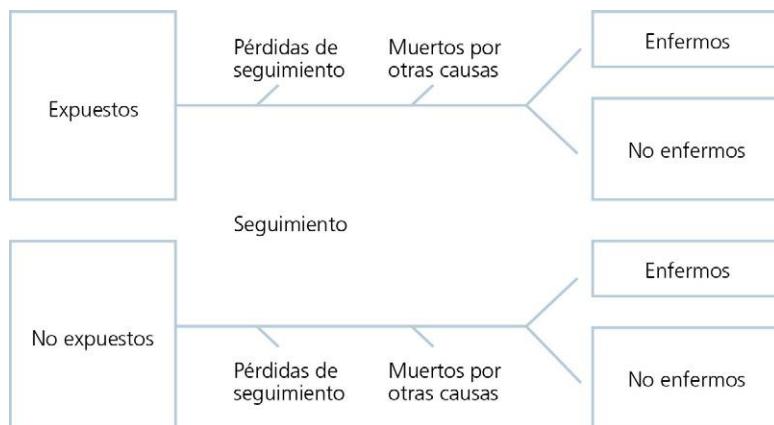
no presentaban enfermedad coronaria. Los miembros de esta cohorte se han seguido durante más de 30 años desde las mediciones iniciales. Se han determinado las tasas de incidencia de enfermedad coronaria en función de las características iniciales, y se han proporcionado evidencias de que los valores elevados de presión arterial y colesterol predicen la enfermedad coronaria, y así fueron considerados como factores etiológicos de la misma.

### TIPOS DE ESTUDIOS DE COHORTES

Según la relación cronológica entre el inicio del estudio y el desarrollo de la enfermedad de interés, los estudios de cohortes se pueden clasificar en prospectivos o retrospectivos.

En un *estudio prospectivo de cohortes*, el investigador parte de la formación de los grupos de sujetos expuestos y no expuestos a un posible factor de riesgo, y los sigue durante un tiempo para determinar las tasas de incidencia del desenlace de interés o de mortalidad en ambos grupos. El carácter prospectivo significa que el investigador recoge la información sobre la exposición cuando se inicia el estudio, e identifica los nuevos casos de la enfermedad o las defunciones que se producen a partir de ese momento, es decir, el inicio del estudio es anterior al desarrollo de los hechos.

En un *estudio retrospectivo de cohortes*, tanto la exposición como la enfermedad ya han ocurrido cuando se lleva a cabo el estudio. La identificación de las cohortes expuesta y no expuesta se basa en su situación en una fecha previa bien definida (p. ej., la fecha de inicio de una exposición laboral),



**Figura 7.1** Estructura básica de un estudio de cohortes.

suficientemente lejos en el tiempo para que la enfermedad en estudio haya tenido tiempo de desarrollarse, y se determina en cuántos sujetos se ha presentado el efecto de interés.

En algunas circunstancias, los estudios pueden ser *ambispectivos*, en los que se recogen datos de forma retrospectiva y prospectiva en una misma cohorte.

**Ejemplo 7.2.** Supongamos que en el año 2000 se decidió realizar un estudio de las consecuencias de la exposición a un agente tóxico. Se seleccionó una cohorte de 1.500 individuos que habían estado expuestos al agente entre 1990 y 1995, y una cohorte de sujetos que no lo habían estado. Se recogieron datos retrospectivos con el fin de determinar si habían existido diferencias entre ambas cohortes en la aparición de problemas de salud, como defectos congénitos, infertilidad y trastornos mentales, entre otros. Además, como existía la posibilidad de que este agente incrementara el riesgo de cáncer, las cohortes se siguieron prospectivamente desde el año 2000 hasta 2010.

Aunque los estudios retrospectivos se pueden llevar a cabo de forma más rápida y económica, ya que tanto la exposición como la enfermedad ya han ocurrido cuando se inicia el estudio, sólo pueden realizarse si existe un mecanismo adecuado para identificar las cohortes y registros completos de estas personas. Si los datos son incompletos o no son comparables para todos los sujetos, el estudio carece de validez. Además, en muchas ocasiones la información sobre factores de confusión no está disponible.

Los estudios de cohortes también pueden clasificarse según si utilizan un grupo de comparación interno o externo. Cuando se estudia una sola cohorte

general (p. ej., los residentes en un área geográfica determinada, como es el caso del estudio de Framingham), pueden distinguirse dos *cohortes internas* como consecuencia de la clasificación de los sujetos de la cohorte general en expuestos y no expuestos.

**Ejemplo 7.3.** Supongamos que se desea estudiar la relación entre la exposición al tabaco y la enfermedad coronaria entre los médicos que ejercen en una comunidad autónoma. La cohorte expuesta la formarían los médicos que fuman y la cohorte no expuesta los médicos no fumadores o que fuman una cantidad inferior que los clasificados como expuestos. Existe una cohorte general definida por la pertenencia a un grupo profesional y un criterio geográfico, cuyos sujetos se dividen en dos o más cohortes internas en función del grado de exposición.

En otras situaciones no existe una cohorte general bien definida, sino que la cohorte expuesta se ha definido por su alta frecuencia de exposición, como, por ejemplo, trabajadores de determinadas industrias. En estos casos se utiliza una cohorte de comparación *externa*, formada por un grupo de sujetos no expuestos al factor de estudio que son seguidos simultáneamente con la cohorte expuesta.

**Ejemplo 7.4.** Supongamos que se desea estudiar si un determinado agente químico aumenta la incidencia de cáncer. La cohorte expuesta la forman trabajadores de una industria expuestos al producto químico, mientras que la cohorte de comparación la podrían formar trabajadores de otra industria no expuestos a dicho producto. En este caso no existe una única cohorte general, sino que las cohortes son grupos de sujetos seleccionados de forma independiente.

En ocasiones, la cohorte externa de comparación puede provenir de la población general, y estaría formada por una muestra de la misma que sería seguida simultáneamente con la cohorte de estudio.

Algunos estudios sólo siguen una cohorte de sujetos expuestos y utilizan, como referencia, información registrada procedente de la población general.

**Ejemplo 7.5.** Si se estudia un grupo de individuos que trabajan en una industria con un riesgo aumentado de desarrollar cáncer de vejiga urinaria, la mortalidad por este tipo de cáncer en estos sujetos puede compararse con las cifras de mortalidad de la población general. En este caso no se sigue una cohorte de sujetos provenientes de la población general, sino tan sólo las estadísticas referidas a ella. Por esta razón en estos estudios las únicas variables de confusión que pueden controlarse son la edad, el sexo u otras que aparezcan en los registros de mortalidad.

## **ESTIMACIONES QUE PUEDEN REALIZARSE EN LOS ESTUDIOS DE COHORTES**

Los estudios de cohortes son longitudinales. Al existir seguimiento de los sujetos, permiten realizar las siguientes estimaciones:

- La incidencia de la enfermedad en los sujetos expuestos y en los no expuestos (anexo 1). Además, en los individuos expuestos permite estimar la incidencia según diferentes características de la exposición: intensidad, diferentes períodos de tiempo, diversas combinaciones de los factores de riesgo, etc. También permite evaluar si los cambios en los valores de exposición se relacionan con la incidencia de la enfermedad.
- El riesgo relativo como medida de la magnitud de la asociación entre el factor de riesgo y la variable de respuesta. Estima el riesgo de los sujetos expuestos de presentar la respuesta en relación con los no expuestos (anexo 2). También permite estimar el riesgo relativo según diferentes características de la exposición, y se puede evaluar, por ejemplo, la existencia de un gradiente dosis-respuesta entre la exposición y la aparición del efecto.
- La fracción o proporción atribuible o proporción de casos de una enfermedad que resulta de la exposición a un factor determinado o a una combinación de ellos (ver capítulo 32).

- La diferencia de incidencias como medida del impacto potencial que tendría la eliminación de la exposición (ver capítulo 32).

## **IDENTIFICACIÓN DE LAS COHORTE**

Para que la duración del estudio, el coste y el tamaño de la muestra sean mínimos hay que seleccionar a individuos que tengan una alta probabilidad de presentar la enfermedad o el desenlace de interés, y excluir a aquellos que ya la presentan o que no puedan desarrollarla. Con frecuencia la identificación de sujetos con antecedentes de la enfermedad o que la padecen de forma asintomática o subclínica puede resultar difícil. Si han de aplicarse a toda la cohorte, las pruebas y exploraciones complementarias que se utilicen para tratar de identificarlos deben ser sencillas, seguras y baratas.

Dado que son estudios analíticos (estudian una presunta relación causa-efecto), la consideración más importante es la comparabilidad de las cohortes expuesta y no expuesta, así como asegurar que se dispone del número suficiente de sujetos con las características predictoras más importantes y de desenlaces observados durante el estudio, a fin de permitir un análisis válido. Cuando se utiliza una cohorte interna, dado que la cohorte no expuesta procede de la misma población que la expuesta, la comparabilidad es más fácil de asegurar. En cambio, cuando la cohorte de comparación es externa, es más difícil asumir que ambas cohortes proceden de la misma población.

### **Identificación de la cohorte expuesta**

La cohorte expuesta puede proceder de la población general o de grupos especiales en los que la exposición es frecuente o en los que resulta fácil efectuar un seguimiento completo.

### **Población general**

En ocasiones se desea que la cohorte sea representativa de la población general de un área geográfica bien definida, como ocurre en el Framingham Heart Study. A partir de esta cohorte general, los individuos se clasifican en expuestos y no expuestos. Este tipo de diseño es útil para investigar la relación entre uno o varios factores de riesgo y una o varias enfermedades frecuentes en la población. Es conveniente seleccionar un área geográfica en la que la población sea estable, con pocos fenómenos migratorios, y en la que se puedan identificar claramente hospitalizaciones.

y centros de salud de referencia donde, con toda probabilidad, acudirá la mayoría de la población cuando enferme, de forma que se facilite el seguimiento de la cohorte.

Este tipo de estudio es deseable por su capacidad de generalización, pero suele ser muy caro y presenta una elevada proporción de pérdidas de seguimiento. Además, a no ser que el estudio se limite a sujetos con un alto riesgo de padecer la enfermedad o el desenlace de interés, el tamaño de la muestra necesario para que aparezca el número suficiente de casos puede ser muy elevado.

### Grupos especiales en los que la exposición es frecuente

Cuando la exposición es rara, las dificultades y el coste que representa estudiar una muestra de la población general son importantes. Por ello es preferible seleccionar un grupo específico de individuos expuestos al factor de estudio y compararlo con una muestra de no expuestos. Con frecuencia, se seleccionan sujetos que trabajan en una determinada industria u ocupación, ya que en ellos la exposición de interés es más frecuente, existe una menor probabilidad de pérdidas de seguimiento, suele disponerse de información relevante en sus registros médicos y laborales, y en muchas circunstancias son sometidos a exámenes médicos periódicos. Estos grupos son fáciles de seguir y permiten obtener un número suficiente de personas expuestas en un período de tiempo razonable. Además, dado que suelen haber estado expuestos a una elevada intensidad del factor de estudio, si existe una relación causa-efecto se necesitarán menos individuos para demostrarla que si se hubiera estudiado una muestra de la población general.

### Grupos especiales que facilitan el seguimiento

En otras situaciones los grupos se seleccionan porque permiten recoger la información que se desea de forma más sencilla y menos costosa, y facilitan el seguimiento (p. ej., médicos, profesionales de enfermería, alumnos universitarios, etc.). Estas ventajas también las presentan las mujeres embarazadas, ya que acuden con frecuencia a los centros sanitarios para el control de su embarazo. Dado que, además, el período de seguimiento necesario es corto, los estudios de cohortes han sido muy utilizados para evaluar el riesgo teratogénico de fármacos u otras exposiciones.

Aunque las ventajas de utilizar cohortes seleccionadas hacen que esta estrategia sea la preferida en muchas ocasiones, siempre que se utilizan cohortes no representativas de la población general se debe

ser cauto a la hora de aplicar los resultados a la población como conjunto.

### Identificación de la cohorte no expuesta

La función de la cohorte de referencia es estimar la frecuencia que tendría la enfermedad en la población de donde provienen los individuos que componen la cohorte expuesta si no existiera exposición. Si la frecuencia es similar en ambos grupos, sugiere que el factor en estudio no aumenta la probabilidad de desarrollar la enfermedad. Para poder efectuar esta afirmación, la cohorte no expuesta debe ser comparable a la expuesta respecto a las variables que se consideran pronósticas de la enfermedad, a excepción del factor que se está estudiando.

La elección del grupo de comparación deriva directamente de la elección del grupo de expuestos. Si se selecciona una cohorte general, en la que los sujetos son clasificados según su exposición, la referencia natural es el grupo formado por los sujetos que no presentan la exposición. A este grupo control o de referencia se le denomina cohorte interna, ya que proviene de la misma cohorte general que se estudiaba. La utilización de una cohorte interna no sólo aumenta la probabilidad de que tanto expuestos como no expuestos procedan de un mismo subgrupo de población, sino que habitualmente implica que también estarán sujetos a los mismos procedimientos de seguimiento y tendrán, por tanto, la misma probabilidad de que se les detecte la enfermedad.

La selección de una cohorte externa suele ser la mejor elección para estudiar exposiciones raras u otras exposiciones a posibles factores de riesgo laborales o ambientales, aunque será más difícil asegurar que las dos cohortes proceden de la misma población. Por esta razón el problema que pueden plantear los posibles factores de confusión se acentúa en estos estudios. Las cohortes pueden ser diferentes en otros aspectos importantes, además de la exposición a la variable de interés, capaces de influir sobre la enfermedad. Algunas de estas diferencias potencialmente generadoras de confusión pueden medirse y llegar a controlarse, pero otras son desconocidas o imposibles de determinar.

Si la cohorte expuesta la componen grupos seleccionados por su alta frecuencia de exposición, como trabajadores de determinadas industrias, se puede elegir como referencia a un grupo de individuos que trabajan en otras industrias y que no estén expuestos al factor de estudio, o bien trabajadores de la misma empresa cuyo puesto de trabajo no implique la exposición al factor de estudio. Los individuos de esta cohorte externa deben ser

semejantes a los de la cohorte expuesta en cuanto a características demográficas y socioeconómicas y otros factores relevantes que puedan influir sobre la enfermedad.

Una alternativa es comparar la frecuencia de desenlaces observada en la cohorte expuesta con los datos procedentes de la población general obtenidos a partir de registros. Para efectuar dicha comparación se ha de disponer de información sobre el desenlace o la enfermedad en estudio referida a toda la población (p. ej., estadísticas de mortalidad, registros de cáncer poblacionales). Se asume que la proporción de expuestos al factor entre la población general es pequeña (o se tiene información sobre su frecuencia), ya que, de otro modo, se infraestimaría la verdadera asociación.

La población general puede diferir de la cohorte expuesta por diferentes factores (p. ej., en los estilos de vida), cuya información no suele estar disponible en los registros de mortalidad, por lo que no puede controlarse su posible efecto de confusión. Si la cohorte expuesta está formada por individuos que trabajan en una industria, debe tenerse en cuenta que las personas que tienen y mantienen un empleo gozan, en general, de mejor salud que los que no trabajan (efecto del trabajador sano). Los sujetos que trabajan suelen tener mayores ingresos económicos y mejor acceso a la atención médica, y pueden tener estilos de vida que redunden en una mejor salud. Dado que la población general incluye a individuos que no trabajan debido a enfermedad o incapacidad, las tasas de mortalidad son casi siempre mayores en la población general. La consecuencia de este fenómeno de selección es que cualquier exceso de riesgo asociado a una ocupación específica se infraestimará al compararla con la población general.

La consideración de que las personas seleccionadas por su exposición pueden estar más o menos sanas que la población general se refiere no sólo a la exposición de interés sino también a otros factores etiológicos. Por ejemplo, si la exposición de interés es el uso de anticonceptivos orales, el problema es decidir si el grupo de comparación deben ser mujeres que no los utilizan o limitarse a mujeres que utilizan otros métodos de anticoncepción, ya que las que no utilizan ningún método pueden ser diferentes de las que sí las toman en relación con sus prácticas sexuales, deseos de quedar embarazadas o situación fértil.

En ocasiones puede considerarse de interés la utilización de más de un grupo de comparación. Si se observan resultados diferentes al comparar con cada uno de ellos, debe preguntarse el porqué. Por otro lado, estas inconsistencias producen una mayor incertidumbre en los resultados.

## Cohorte fija o cohorte dinámica

Se habla de *cohorte fija* cuando sólo se incluyen los individuos que cumplen los criterios de inclusión en la fecha de inicio del estudio. Por tanto, la fecha de inclusión es la misma para todos los sujetos y coincide con la del inicio del estudio. Se habla de *cohorte dinámica* cuando los individuos se van incluyendo a medida que progresá el estudio, es decir, a medida que se van identificando. Por tanto, en una cohorte dinámica la fecha de inclusión es diferente para cada sujeto.

Es más probable que exista un sesgo de selección en una cohorte fija, ya que los individuos con un mayor grado de exposición pueden haber desarrollado la enfermedad y, por tanto, quedarían excluidos de la cohorte inicial. En cambio, en una cohorte dinámica pueden incluirse los sujetos en un mismo momento de su historia de exposición.

## MEDICIÓN DE LA EXPOSICIÓN

Uno de los puntos fuertes de los estudios de cohortes prospectivos es que permiten obtener una información detallada, precisa y objetiva de la exposición en estudio. Así, por ejemplo, si se desea estudiar el efecto de determinados hábitos alimentarios sobre la incidencia de enfermedades cardiovasculares o diferentes tipos de cáncer, el grado de detalle necesario sobre la alimentación hace que la recogida retrospectiva de esta información sea poco fiable.

La medición detallada de la exposición permite estudiar el efecto en un subgrupo de sujetos sometidos a determinado grado o subtipo de exposición, o incluso clasificar la exposición en diferentes grados para evaluar una posible relación dosis-respuesta.

La duración de la exposición puede utilizarse, en ocasiones, como un sustituto de la dosis acumulada de exposición, siempre que la intensidad de la exposición sea poco variable en el tiempo.

Por otro lado, el grado de exposición puede sufrir cambios durante el seguimiento, por lo que es importante anotar tanto los cambios como sus motivos.

## MEDICIÓN DEL DESENLACE

Los procedimientos para determinar qué sujetos de las cohortes desarrollan el desenlace de interés varían en función de la respuesta que se estudia y los recursos disponibles. Esta medición debe ser lo más objetiva posible. Cuando se trata de síntomas o síndromes para los que no existe ningún examen objetivo, debe recurrirse a definiciones sobre las que se disponga de un amplio consenso.

Es importante que esta determinación se realice de la misma forma en los expuestos y los no expuestos, mediante criterios establecidos explícitamente antes del inicio del estudio. Además, en la medida de lo posible, esta medición debería ser realizada por personas que no conozcan si los sujetos están expuestos o no, para prevenir sesgos de información.

En muchos estudios el resultado de interés es si la persona fallece y, en caso afirmativo, cuáles han sido las causas específicas. La utilización de la mortalidad como variable de respuesta sólo es razonable si la enfermedad de interés es frecuentemente letal. A pesar de sus limitaciones, en general, se acepta como causa de muerte la que aparece en el certificado de defunción, pero en otros casos se recoge información adicional procedente de registros médicos, autopsias, etc., con la finalidad de aumentar su calidad.

Cuando las personas que presentan la respuesta suelen requerir hospitalización, pueden monitorizarse los registros de los hospitales de referencia para establecer su aparición. A menudo se utilizan los informes de alta por la rapidez y bajo coste de su uso. Sin embargo, dado que no puede asumirse su exactitud ni su exhaustividad, suele ser necesario recurrir a la historia clínica para complementar la información. Los procedimientos para combinar información procedente de diferentes fuentes deben establecerse claramente a priori.

Cuando la información procede de registros ya existentes, pueden presentarse problemas relacionados con la ausencia de estandarización de la información, la variabilidad de los criterios diagnósticos utilizados o la exhaustividad de la información contenida en los registros. Además, los sujetos pueden solicitar atención a centros o médicos diferentes de los monitorizados, por lo que puede sobreestimarse o infraestimarse la incidencia de la enfermedad. También pueden utilizarse registros poblacionales, si existen. Por estas razones es conveniente examinar de forma periódica a los miembros de la cohorte, aunque ello suponga un coste importante.

Cuando el diagnóstico requiere la realización de examen físico y exploraciones complementarias, es preferible utilizar criterios estándar reconocidos, lo que permitirá comparar los resultados con los de otros estudios. En algunos casos puede resultar útil subdividir el diagnóstico según su grado de certeza o su etiología. La combinación de todos los subtipos como un solo grupo dificulta la identificación de los factores de riesgo.

También debe tenerse en cuenta que la enfermedad puede ser indetectable clínicamente, tanto si es infecciosa (hepatitis, poliomielitis, etc.) como si no lo es (arteriosclerosis coronaria, cáncer, etc.).

Los factores de riesgo pueden tener una influencia diferente en los casos clínicos y en los subclínicos, posibilidad que deberá tenerse en cuenta en el análisis. No debe olvidarse que la existencia de resultados falsos positivos y falsos negativos puede sesgar los resultados.

## SEGUIMIENTO

El mayor desafío de un estudio de cohortes es el seguimiento de un elevado número de sujetos durante un período prolongado de tiempo, por lo que la mayoría de los esfuerzos se dedican a asegurarlo. Los investigadores deben recoger información sobre los cambios que se presentan en los factores de riesgo y las variables pronósticas, así como registrar si alguno de los participantes ha desarrollado el desenlace de interés. Algunos individuos modificarán su exposición durante el seguimiento. Por ejemplo, si el factor en estudio es el hábito tabáquico, algunos sujetos empezarán a fumar durante el estudio y otros que fumaban al inicio dejarán de hacerlo.

El seguimiento debe ser el mismo en todos los grupos, con idéntica pauta de visitas y exploraciones. Así se intentan evitar los efectos derivados de un mayor control en uno de los grupos, lo que podría influir tanto en el sujeto que se sabe estudiado como en el investigador y su capacidad u oportunidad de detectar sucesos que, de otra manera, pasarían inadvertidos. Este control se facilita con las técnicas de enmascaramiento, si es posible aplicarlas.

Deben preverse los sucesos o situaciones que obligarán a la exclusión de un sujeto y las causas de posibles pérdidas durante el seguimiento, sobre todo en estudios de larga duración. Hay que preparar estrategias para minimizarlas y, si se producen, recaptarlas o, como mínimo, anotar los motivos por los que se han producido.

El tiempo de seguimiento debe ser lo suficientemente largo para permitir que aparezca el número suficiente de casos para conseguir la potencia estadística deseada, pero lo suficientemente corto como para no crear graves problemas logísticos y de organización. Si el objetivo es, por ejemplo, conocer si el tabaco es causa de cardiopatía isquémica, 3 meses de seguimiento sería, sin lugar a dudas, un tiempo demasiado corto para observar algún efecto, pero, por otro lado, un seguimiento de varias décadas puede ser demasiado largo como para que los investigadores puedan mantener el control de la cohorte de manera eficaz.

El intervalo entre las distintas visitas de seguimiento depende de la enfermedad y debe ser lo suficientemente corto como para detectar su aparición. Las visitas muy poco espaciadas en el tiempo pueden ser

una molestia para los sujetos y provocar cansancio y falta de interés, que pueden traducirse en pérdidas de seguimiento. Por otro lado, si entre visita y visita transcurre mucho tiempo, el resultado puede ser idéntico.

### Estrategias para reducir las pérdidas

Para reducir el número de pérdidas de seguimiento suele ser útil, al establecer los criterios de selección, excluir a los individuos que es probable que se pierdan por presentar determinados problemas de salud, porque piensan cambiar de domicilio o por cualquier otra razón. Además, al comenzar el estudio debe recogerse información que permita localizar a los individuos en caso de que éstos cambien de domicilio o fallezcan.

Los contactos periódicos con los individuos ayudan a mantener el seguimiento, y también pueden ser útiles para determinar con mayor exactitud el momento en que aparecen los desenlaces de interés. Estos contactos pueden realizarse tanto telefónicamente (mediante intentos repetidos, con llamadas por la noche y durante los fines de semana, si es necesario) como por correo electrónico o postal (p. ej., envíos repetidos en los que se adjuntarán tarjetas o sobres con el franqueo pagado y la dirección impresa).

Para los sujetos con los que no se pueda contactar por teléfono o por correo, deben utilizarse otros procedimientos, como contactar con amigos, familiares o su médico de cabecera, y determinar su situación mediante registros como el de mortalidad o altas hospitalarias, realizar visitas personales, etc. La búsqueda de los individuos para realizar valoraciones de seguimiento requiere, en ocasiones, perseverancia y repetidos esfuerzos por múltiples mecanismos.

## VENTAJAS Y LIMITACIONES DE LOS ESTUDIOS DE COHORTE

En los estudios de cohortes se pueden calcular las tasas de incidencia y el riesgo relativo, lo que supone una ventaja en relación con los estudios de casos y controles, que tan sólo permiten estimar el riesgo relativo de forma aproximada pero no calcular las tasas de incidencia.

La estructura de estos estudios asegura una adecuada secuencia temporal, ya que se sigue a los sujetos desde la exposición hasta la aparición del efecto o desenlace. Esta característica también los diferencia de los estudios de casos y controles, en los que no siempre puede establecerse con seguridad una adecuada secuencia temporal.

La medición de la exposición y de las variables pronósticas tiende a ser más exacta y completa que

en los estudios de casos y controles, ya que no depende, en la mayoría de las ocasiones, de la memoria del paciente ni está influida por la presencia de la enfermedad, en especial cuando la información se recoge de forma prospectiva. En los estudios retrospectivos la calidad de la información depende de los registros y de si éstos se adecuan a los objetivos del estudio.

Los estudios de cohortes permiten evaluar los efectos del factor de riesgo sobre varias enfermedades, por ejemplo de la hipertensión arterial en relación con diferentes desenlaces cardiovasculares.

Los estudios de cohortes no son eficientes para el estudio de enfermedades poco frecuentes, ya que obligan a seguir a un número elevado de sujetos durante un largo período de tiempo para observar un número suficiente de casos de la enfermedad. De todos modos, aunque la frecuencia de una enfermedad sea muy baja en la población general, puede ser lo suficientemente elevada en grupos especiales que hayan estado expuestos de forma intensa a un factor de riesgo y, en consecuencia, puede llevarse a cabo un estudio de cohortes en este grupo de población expuesto. Por ejemplo, aunque el mesotelioma es una enfermedad muy poco frecuente entre la población general, es relativamente común entre los trabajadores de los astilleros, por lo que sería factible realizar un estudio de cohortes con estos sujetos.

**Ejemplo 7.6.** Supongamos un estudio cuyo desenlace de interés es la enfermedad coronaria. Se estima que su incidencia en mujeres de 45 años que presentan bajo riesgo (presión arterial sistólica inferior a 110 mmHg, colesterolemia inferior a 185 mg/dl, sin hipertrofia ventricular izquierda, no fumadoras y sin intolerancia a la glucosa) es del 0,8% a los 6 años. Esto significa que para encontrar 10 casos se debería realizar el seguimiento de unas 1.250 mujeres con estas características durante 6 años.

Si, por el contrario, se selecciona un grupo de hombres de la misma edad con alto riesgo (presión arterial sistólica por encima de 180 mmHg, colesterolemia superior a 310 mg/dl, fumadores, con hipertrofia ventricular izquierda y, además, intolerancia a la glucosa), la incidencia estimada es del 35,8%, lo que implica que sólo será necesario realizar el seguimiento de 28 sujetos con estas características durante el mismo período de tiempo para encontrar el mismo número de casos (Wilson et al, 1987).

Los estudios de cohortes prospectivos tampoco son eficientes para el estudio de enfermedades con un largo período de latencia, ya que obligan a un seguimiento muy largo, lo cual, además del coste y esfuerzo que supone, puede aumentar las pérdidas.

Los estudios de cohortes requieren un número elevado de participantes y tienen un coste elevado. Por estas razones, en general, sólo se llevan a cabo después de que una hipótesis etiológica haya sido evaluada mediante un diseño de casos y controles.

**Ejemplo 7.7.** Un caso es el estudio sobre la utilización de benzodiazepinas durante el primer trimestre de embarazo y la aparición del labio leporino y hendidura palatina en el recién nacido. Existe controversia acerca del efecto que ejerce la ingestión de estos fármacos sobre el aumento de hendiduras orales. Parte de los resultados contradictorios nacen a raíz de que el diseño más utilizado para evaluar esta hipótesis ha sido el de casos y controles, en el que la información sobre la exposición se consigue de forma retrospectiva, después del nacimiento. En este caso, es muy posible que las madres que hayan tenido un niño con fisura palatina tiendan a recordar con más exactitud su historia farmacológica que las que no han padecido este problema. En los estudios de cohortes, al obtener la información sobre los medicamentos antes de que la madre dé a luz se evita este posible sesgo.

Siguiendo este razonamiento, Shiono y Mills (1984) llevaron a cabo un estudio de cohortes. El problema con el que toparon los autores fue que, después de seguir a 854 mujeres expuestas al fármaco, sólo una de ellas tuvo un recién nacido con la malformación. En el grupo no expuesto de 32.364 mujeres, sólo 32 recién nacidos presentaron fisura palatina. Aunque es cierto que no se produjo ningún sesgo de memoria, no pudieron encontrar un resultado estadísticamente significativo debido a que el tamaño de la muestra era insuficiente. Incluso si el riesgo relativo fuera de 3,0, la potencia de este estudio para encontrar esta diferencia sería de un 50%.

Las principales ventajas y desventajas de los estudios de cohortes se resumen en el cuadro 7.1.

## ESTUDIOS RETROSPECTIVOS DE COHORTES

En los estudios retrospectivos de cohortes, también llamados estudios de cohortes históricas, el investigador identifica una cohorte de sujetos según sus características en el pasado, y reconstruye su experiencia de enfermedad hasta un momento definido de tiempo en el pasado más reciente o en el momento actual.

La estructura de un estudio retrospectivo de cohortes es similar a la de un estudio prospectivo: se efectúa un seguimiento de un grupo de individuos a lo largo del tiempo, en los que se miden las posibles variables predictoras al inicio, y después se determinan

Cuadro 7.1 **Ventajas y desventajas de los estudios de cohortes**

### Ventajas

- Permiten el cálculo directo de las tasas de incidencia en las cohortes expuesta y no expuesta, y del riesgo relativo de los expuestos en relación con los no expuestos
- Aseguran una adecuada secuencia temporal (la exposición al factor de estudio precede a la aparición del desenlace)
- En los diseños prospectivos se minimizan los errores en la medición de la exposición
- Permiten evaluar los efectos del factor de riesgo sobre varias enfermedades

### Desventajas

- No son eficientes para el estudio de enfermedades poco frecuentes
- No son eficientes para el estudio de enfermedades con un largo período de latencia
- Los estudios prospectivos suelen ser de larga duración
- Requieren un número elevado de participantes
- Los estudios prospectivos tienen un coste elevado

los desenlaces que se producen. Por tanto, ambos diseños permiten asegurar que la exposición (factor de riesgo) precede a la enfermedad o efecto. La diferencia es que la identificación de la cohorte, las mediciones basales, el seguimiento y los desenlaces ya se han producido cuando se inicia el estudio.

Un uso frecuente de este tipo de estudios es la evaluación de riesgos laborales, aunque se han utilizado con múltiples finalidades.

**Ejemplo 7.8.** Para evaluar la hipótesis de que la pérdida del empleo conlleva una mayor morbilidad, Beale y Nethercott (1986) realizaron un estudio retrospectivo de cohortes. Obtuvieron un listado con el nombre y la dirección de todos los empleados de una fábrica despedidos a causa de su cierre. La compañía proporcionó, además, la siguiente información: fecha de entrada en la fábrica, horas de trabajo, tipo de ocupación y el departamento en el que estaba destinado cada uno de los empleados. Para ser incluidos en el estudio debían cumplir los siguientes requisitos: estar registrado con algún médico de la comunidad donde se realizaba el estudio, haber trabajado como mínimo durante 4 años antes del despido en la misma fábrica y tener una jornada laboral superior a las 37 horas. Se excluyó a los individuos con una edad superior a 61 años debido a que su jubilación estaba próxima.

Los investigadores obtuvieron de las historias clínicas el número de consultas de cada trabajador durante los 4 años previos al anuncio del cierre de la fábrica y las visitas realizadas después del despido. En ningún momento se entrevistó o se recogió información personal. Como grupo control se eligieron trabajadores de otras fábricas con características similares, que cumplían los mismos criterios de inclusión, residían en la misma comunidad y no habían sido despedidos.

En relación con los estudios prospectivos, presentan algunas ventajas relacionadas con el menor tiempo necesario para su realización, y su consiguiente menor coste.

En relación con los estudios de casos y controles, comparten la ventaja de no necesitar un largo período de tiempo para identificar un número suficiente de casos, pero se diferencian en que los sujetos son seleccionados por su exposición. Por tanto, si se estudian exposiciones muy poco frecuentes, este diseño puede ser una forma práctica de asegurar que el número de sujetos expuestos es suficiente, aunque requiere que esté disponible la información sobre la exposición previa. La ausencia de este tipo de información limita el número de situaciones en

que este diseño es factible. La principal limitación de estos estudios es que los investigadores carecen de control sobre la calidad de la información.

Cuando la exposición ha tenido lugar muchos años antes del inicio del estudio, su medición puede ser muy difícil y sólo puede realizarse una clasificación en grandes grupos (sin poder diferenciar los grados de exposición), por lo que es probable que aparezcan errores de medición que sesguen la aparente magnitud de la asociación. Además, al no poder utilizar una clasificación por diferentes grados de exposición, no puede evaluarse la existencia de una relación dosis-respuesta, que es un importante criterio para inferir causalidad.

La naturaleza retrospectiva de la recogida de datos también dificulta la medida de la aparición de la enfermedad. Dado que el investigador debe utilizar información recogida con otras finalidades, su calidad puede no ser suficiente, lo que limita la realización de este tipo de estudios. Además, pueden existir problemas adicionales en la recogida de información sobre importantes variables de confusión. Por tanto, los resultados de muchos estudios retrospectivos difícilmente pueden considerarse definitivos por abordar relaciones causales.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

Beale N, Nethercott S. Job loss and morbidity in a group of employees nearing retirement age. *J R Col Gen Pract*. 1986;36:265-6.

Shiono PH, Mills JL. Oral clefts and diazepam use during pregnancy. *N Engl J Med*. 1984;311:919-20.  
Wilson P, Castelli W, Kannel W. Coronary risk prediction

in adults (The Framingham Heart Study). *Am J Cardiol*. 1987;59:91G-4G.

## BIBLIOGRAFÍA

Breslow NE, Day NE. Statistical methods in cancer research. Vol. II. The design and analysis of cohort studies. Lyon: IARC Publications; 1987.  
Dekkers OM. On causation in therapeutic research: observational studies, randomised experiments and instrumental variable analysis. *Prev Med*. 2011;53(4-5):239-41.  
DiPietro NA. Methods in epidemiology: observational study designs. *Pharmacotherapy*. 2010;30:973-84.  
Euser AM, Zoccali C, Jager KJ, Dekker FW. Cohort studies:

prospective versus retrospective. *Nephron Clin Pract*. 2009;113(3):c214-7.  
Fernández E. Estudios observacionales. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo científico. Barcelona: Elsevier; 2010. p. 205-14.  
Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60(7):578-86.

Manchikanti L, Singh V, Smith HS, Hirsch JA. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: part 4: observational studies. *Pain Physician*. 2009;12(1):73-108.  
Samet JM, Muñoz A. Evolution of the cohort study. *Epidemiol Rev*. 1998;20(1):1-14.  
Samet JM, Muñoz A. Perspective: cohort studies. *Epidemiol Rev*. 1998;20(1):135-6.  
Silverman SL. From randomized controlled trials to observational studies. *Am J Med*. 2009;122(2):114-20.

- Shields KM, DiPietro NA, Kier KL. Principles of drug literature evaluation for observational study designs. *Pharmacotherapy*. 2011;31(2):115-27.
- Szklo M. Population-based cohort studies. *Epidemiol Rev.* 1998;20(1):81-90.
- Tager IB. Outcomes of cohort studies. *Epidemiol Rev.* 1998;20(1):15-28.
- Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med.* 2007;4(10):e297.
- Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet.* 2007;370(9596):1453-7.
- White E, Hunt JR, Casso D. Exposure measurement in cohort studies: the challenges of prospective data collection. *Epidemiol Rev.* 1998;20(1):43-56.
- Whitney CW, Lind BK, Wahl PW. Quality assurance and quality control in longitudinal studies. *Epidemiol Rev.* 1998;20(1):71-80.
- Willett WC, Colditz GA. Approaches for conducting large cohort studies. *Epidemiol Rev.* 1998;20(1):91-9.
- Yang W, Zilov A, Soewondo P, Bech OM, Sekkal F, Home PD. Observational studies: going beyond the boundaries of randomized controlled trials. *Diabetes Res Clin Pract.* 2010;88(Suppl 1):S3-9.

## Estudios de casos y controles

En los estudios de casos y controles se elige un grupo de individuos que tienen un efecto o una enfermedad determinada (casos), y otro en el que está ausente (controles). Ambos grupos se comparan respecto a la frecuencia de exposición previa a un factor de riesgo (factor de estudio) que se sospecha que está relacionado con dicho efecto o enfermedad. La estructura básica se recoge en la figura 8.1.

**Ejemplo 8.1.** Si se quiere evaluar el riesgo de padecer un fibroadenoma de mama asociado al uso de anticonceptivos orales (AO), el grupo de casos lo formarían mujeres diagnosticadas de fibroadenoma, y el de control, mujeres que no tuvieran dicha enfermedad, y se debería comparar la historia previa de utilización de AO en ambos grupos.

La necesidad de un grupo control es evidente. El hallazgo, por ejemplo, de que el 47% de las mujeres diagnosticadas de fibroadenoma ha utilizado AO suscita de inmediato la siguiente pregunta: ¿esta proporción de exposición es superior, igual o inferior a la esperada? La función del grupo control es estimar la proporción de exposición esperada en un grupo que no padece la enfermedad.

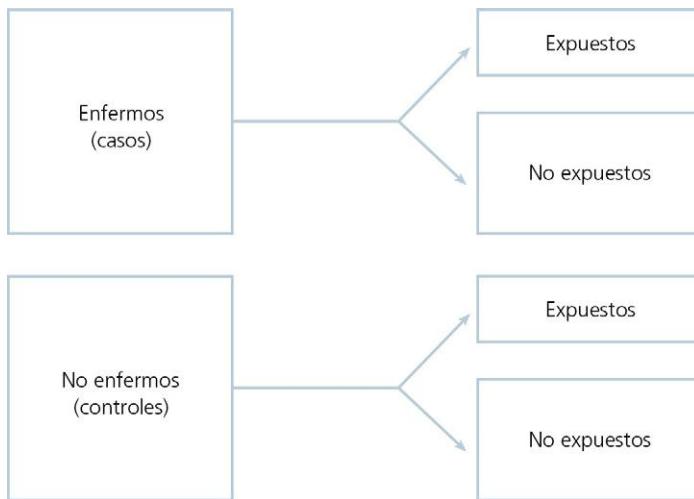
Las estimaciones que se obtienen son la proporción de casos y controles expuestos a un posible factor de riesgo. También son de interés la intensidad y la duración de la exposición en cada uno de los grupos. La medida de asociación o del riesgo de padecer un determinado problema de salud asociado a la presencia de una exposición es la *odds ratio* (OR) (anexo 2).

### SELECCIÓN DE LOS CASOS

Para identificar los casos deben establecerse, de forma clara y explícita, la definición de la enfermedad y los criterios que deben cumplir aquellos que la presenten para ser incluidos en el estudio. Por otro lado, los criterios de selección deben dirigirse a que sólo se incluyan sujetos que potencialmente han podido estar expuestos al presunto factor de riesgo. Estos criterios deben aplicarse por igual a casos y a controles.

**Ejemplo 8.2.** En el estudio hipotético del ejemplo 8.1 se excluiría a las mujeres que presentaran alguna contraindicación para la toma de AO. Las mujeres diabéticas, hipertensas o dislipémicas no entrarían a formar parte del estudio, ya que en estas enfermedades el uso de AO está contraindicado, o al menos su indicación es muy individualizada. Los criterios de exclusión dependen de la definición de exposición que se adopte. Si el interés se centra en el uso de AO en los últimos 5 años o más, una mujer recién diagnosticada de diabetes entraría a formar parte del grupo de casos por tener la probabilidad de haber estado expuesta al factor de riesgo.

Lo ideal sería que los casos seleccionados fueran una muestra aleatoria de todos los individuos de la población de estudio que presentan la enfermedad. En los estudios de cohortes la enfermedad se busca de forma sistemática en todos los participantes, pero en los estudios de casos y controles los casos se obtienen de una serie de pacientes en quienes la enfermedad ya ha sido diagnosticada y están



**Figura 8.1** Estructura básica de un estudio de casos y controles.

disponibles para el estudio, por lo que pueden no ser representativos de la población de estudio.

### Definición de caso

Los investigadores pueden estudiar formas leves y/o graves de la enfermedad. Si se incluyen casos con todo el espectro de gravedad, existe el riesgo de clasificar mal a individuos sin la enfermedad como casos leves, y viceversa, dado que en muchas enfermedades es muy difícil conseguir un diagnóstico de certeza. Si se incluyen sólo formas graves, el riesgo de clasificar mal disminuye, aunque obtener el número de individuos necesario puede ser difícil y, además, se limita la generalización de los resultados.

En algunos casos existe desacuerdo entre distintos investigadores en la definición de una enfermedad, lo que puede introducir variabilidad en la estimación del efecto y reducir la posibilidad de reproducir los resultados. Siempre que exista una definición estándar y ésta sea adecuada para el propósito del estudio, debe usarse con el fin de disminuir la variabilidad y de que los hallazgos sean comparables con los de otros investigadores.

**Ejemplo 8.3.** El síndrome del túnel carpiano puede presentarse con sintomatología diversa, no siempre fácil de diferenciar por criterios únicamente clínicos, de las acroparestesias nocturnas. Los signos de Tinel y Phalen positivos pueden ser de gran ayuda al reproducir la sintomatología, pero sólo tienen valor de sospecha diagnóstica. Según la gravedad del cuadro, existe una mayor o menor afectación sensitiva o motora, no siempre objetivable a la exploración

física. Por esto, para conseguir un diagnóstico preciso y fiable de la afectación del nervio mediano, es preferible utilizar pruebas electrofisiológicas, y no conformarse con los criterios clínicos.

La elección de unos criterios más o menos estrictos para definir la enfermedad tiene implicaciones en los resultados del estudio. Si se aplican criterios diagnósticos poco sensibles y específicos, se hallarán muchos falsos positivos y negativos, lo que tiende a diluir el efecto que se está estudiando. En el caso de que no sea posible aplicar en todos los casos las mejores técnicas para el diagnóstico, es conveniente clasificarlos en categorías como «probable» o «definitivo». Es de esperar que en el grupo con diagnóstico «definitivo» se encuentren menos falsos positivos que en el de «probable», lo que permitirá analizar con mayor detenimiento los posibles errores de una mala clasificación.

### Identificación de los casos

Una vez se ha elaborado una definición conceptual del caso, el siguiente paso es desarrollar una definición operativa para identificarlos. El objetivo es conseguir una identificación en la que todos los casos tengan la misma probabilidad de ser incluidos en el estudio y que sea válida, es decir, que no se incluya ningún individuo que no pertenezca al caso.

**Ejemplo 8.4.** Un ejemplo podría ser la definición conceptual de caso infectado por el VIH (una persona que tendría un resultado positivo si se le practica la prueba) frente a la definición operativa

(una persona a quien se ha practicado la prueba y ha dado un resultado positivo). Los factores que influyen en la decisión de realizar la prueba y la accesibilidad a la misma, al igual que su sensibilidad y especificidad, determinarán quién se considerará caso en la definición operativa.

Otro aspecto importante es el momento en la historia de la enfermedad en que se diagnostican los casos. Siempre que sea posible, es preferible incluir sólo casos recién diagnosticados (casos incidentes). Los casos prevalentes pueden ser los supervivientes, y la supervivencia puede estar relacionada con la exposición. En este caso, se hallaría un riesgo menor que al analizar casos incidentes. En el otro extremo, si la variable en estudio es un factor de buen pronóstico, la inclusión de casos prevalentes llevaría a la sobreestimación de la asociación.

**Ejemplo 8.5.** Supongamos que se quiere evaluar el riesgo de padecer un accidente cerebrovascular (ACV) asociado al consumo de tabaco. Es probable que la proporción de expuestos entre los casos más graves y que mueren a las pocas horas sea mayor que entre los que sobreviven. De hecho, en un estudio se encontró que el riesgo de ACV (letal y no letal) en mujeres de mediana edad fumadoras de 25 cigarrillos o más era del 3,7%. La relación tabaco-ACV era más notable en los casos letales, concretamente de 6 en las fumadoras de más de 25 cigarrillos (Colditz et al., 1988). Así pues, si se estudiaran sólo casos prevalentes (no letales) se hallaría un riesgo menor que al analizar casos incidentes (casos nuevos, tanto letales como no letales).

Si no es viable seleccionar únicamente casos incidentes, pueden incluirse casos prevalentes que hayan sido diagnosticados poco tiempo antes de haber iniciado el estudio. Para evaluar si se ha introducido un sesgo, se pueden estratificar los casos según la fecha del diagnóstico y analizar las frecuencias de exposición en cada estrato y observar si existen diferencias entre ellas.

## SELECCIÓN DE LOS CONTROLES

La elección del grupo control es la clave de este tipo de estudios, ya que debe ser comparable al de casos (los controles deben tener la misma probabilidad de haber estado expuestos que los casos). En realidad, lo importante es que los controles sean representativos de la población de donde provienen los casos.

Conceptualmente, esta población la forman los miembros de una cohorte subyacente determinada por unos criterios de inclusión y exclusión definidos por el

investigador, que será la fuente de los casos. Idealmente, los controles deberían proceder de esta misma población. Un tipo de estudio paradigmático en este sentido es el de *casos y controles anidado en una cohorte* (que se comenta más adelante), en el que se puede identificar claramente la cohorte subyacente, de la que proceden tanto los casos como los controles.

En algunos casos, la población de referencia está bien definida temporal y geográficamente. Un ejemplo serían los estudios de casos y controles poblacionales, en los que se seleccionan todos los casos que se desarrollan en una zona determinada durante un tiempo pre establecido. Esto es factible cuando se trata de enfermedades que requieren atención médica y se dispone de registros exhaustivos. Dado que los casos obtenidos a partir de los registros de enfermedad suelen ser representativos de la población general de pacientes con la enfermedad que habitan en el área, el grupo control debe ser una muestra representativa de la población que reside en el área cubierta por el registro. En estas circunstancias, la selección del grupo control se simplifica, ya que puede extraerse una muestra aleatoria de la población mediante técnicas probabilísticas.

Cuando la detección de todos los casos de una población no es posible, no puede definirse temporal y geográficamente la cohorte subyacente de donde proceden. Esto sucede con frecuencia cuando se trata de enfermedades con síntomas leves que no llevan al individuo a solicitar atención médica, como, por ejemplo, en un estudio sobre las causas de la infertilidad en los hombres, ya que ésta sólo se habría detectado en aquellos que hubieran tratado de tener descendencia y buscado atención médica. En esta situación, los controles deberían ser representativos de aquellos individuos que, en el supuesto de desarrollar la enfermedad, hubieran sido detectados como casos. Si no se restringe la población de acuerdo con estos supuestos, se podría cometer un sesgo de selección relacionado con alguna variable ligada al hecho de buscar atención médica.

**Ejemplo 8.6.** En un estudio de base hospitalaria, los casos podrían ser todos los pacientes diagnosticados en uno o varios hospitales. Los controles serían todos aquellos individuos que en el supuesto de desarrollar la enfermedad serían diagnosticados y, por tanto, tendrían la posibilidad de considerarse casos en aquellos hospitales. La dificultad es, desde luego, identificar con exactitud quién hubiera sido diagnosticado en estos hospitales si hubiera desarrollado la enfermedad.

Para asegurar la comparabilidad de los grupos, debe conseguirse que los controles sean representativos

de la población de referencia de donde proceden los casos, obtenerse información válida tanto de los casos como de los controles, y controlarse los posibles factores de confusión.

### Criterios de selección de los controles

Los controles deben ser comparables a los casos en el sentido de haber tenido la misma probabilidad de exposición. En la práctica, hay que excluir del grupo control a los pacientes que tengan enfermedades relacionadas positiva o negativamente con la exposición en estudio. Si se quisiera estudiar el posible papel protector de la aspirina en la enfermedad coronaria, se debería excluir como controles, por ejemplo, a los pacientes con enfermedades reumáticas (pueden tomar aspirina como parte de su tratamiento) o aquellos con enfermedades gástricas (que probablemente evitarán su uso).

Igualmente, con el fin de evitar que los controles puedan concentrarse en alguna enfermedad relacionada con la exposición, conviene escoger controles con distintas enfermedades con el fin de minimizar un posible sesgo debido al muestreo.

También suele excluirse a los sujetos a los que es muy difícil localizar, de los que se sospecha que no participarán en el estudio, así como los que producen una falta de variabilidad en la exposición (p. ej., en un estudio sobre los efectos de los anticonceptivos se puede excluir a mujeres que, por su edad, muy probablemente no los habrán usado). Si existen evidencias de que un efecto se da principalmente en un subgrupo determinado de la población, la exclusión de otros subgrupos será la mejor estrategia para demostrar su existencia. Si la magnitud del efecto en este subgrupo es importante, la potencia estadística del estudio puede ser mayor si el estudio se restringe a él que si se escoge una población más amplia, a pesar de que el tamaño de la muestra sea más reducido.

### Procedencia de los controles

Los controles pueden seleccionarse entre los sujetos que acuden a los centros sanitarios, directamente de la comunidad o, en algunos casos, se puede recurrir a grupos especiales como vecinos, amigos o familiares de los casos. La decisión de elegir una u otra fuente depende de la población de donde procedan los casos y de las ventajas, y desventajas, de cada una de las fuentes.

La selección de controles a partir de centros de salud u hospitales tiene diversas ventajas: su identificación es mucho más sencilla, el hecho de que los controles tengan alguna enfermedad por la que están siendo visitados hace que puedan recordar con más detalle sus antecedentes personales y la historia de su

possible exposición, y tienden a cooperar más que los individuos sanos, de modo que se consigue reducir la falta de respuestas. Su principal desventaja es que, al padecer algún tipo de enfermedad, es probable que difieran de los individuos sanos en una serie de factores que tienen relación con el proceso de enfermar en general, como el consumo de tabaco o alcohol o la dieta, y que pueden estar relacionados de forma directa o indirecta con la exposición. Por esto, cuando se utilizan este tipo de controles es conveniente incluir a pacientes con distintos diagnósticos con la finalidad de diluir este riesgo.

La utilización de individuos de la población general como controles evita este último problema. Sin embargo, su elección entraña dificultades de orden práctico y metodológico. En primer lugar, entrevistar a individuos sanos es mucho más costoso en tiempo y dinero. En segundo lugar, debido a que muchos de ellos son individuos que trabajan y que tienen su tiempo libre ocupado, no serán tan cooperativos como los pacientes hospitalizados o que acuden a los centros de salud. Las preguntas no contestadas constituirán un problema en la medida en que los individuos que aceptan participar y los que no, difieren por alguna característica relevante para el estudio. Por ejemplo, entre los que aceptan puede haber una proporción más elevada de individuos sin empleo, o de baja laboral, con más tiempo libre para dedicarlo a contestar unas preguntas que entre los que no aceptan. Si el hecho de no trabajar está relacionado de forma directa o indirecta con algún factor de estudio, un elevado porcentaje de preguntas no contestadas puede sesgar los resultados. En tercer lugar, debido a que los controles poblacionales por lo general están menos motivados por el estudio, tenderán a recordar con menos precisión su historia pasada. Ello puede dar lugar a que la calidad de la información que se recoge sea diferente para los casos y los controles. En la tabla 8.1 se resumen las ventajas y desventajas de la utilización de controles procedentes de la población general o de sujetos demandantes.

Una tercera fuente de controles la proporcionan personas relacionadas con los casos, como parientes, amigos o vecinos. Este grupo presenta la ventaja de reducir los costes de realización de la entrevista, en particular en tiempo de desplazamiento, y ofrece más garantías de comparabilidad, al tener más probabilidad de que provengan de la misma base poblacional que los casos. Además, tenderán a ser mucho más colaboradores que la población general por el interés que despierta la cercanía del caso. El inconveniente es el potencial riesgo de sobreemparejamiento que conlleva, dado que algunos hábitos de vida son compartidos por familiares o amigos, e incluso tienden a concentrarse en los mismos barrios.

**Tabla 8.1 Ventajas y desventajas de controles de la población general o la población demandante**

Población demandante	Población general
• Más fáciles de identificar	• Más tiempo y dinero para su identificación
• Tienden a recordar mejor su historia de exposición	• Más probabilidad de que se produzca un sesgo de memoria
• Más cooperativos, por lo que el número de no respuestas es menor	• Tienden a ser menos cooperativos
• Al estar enfermos es más fácil que tengan algún factor de riesgo relacionado con la enfermedad de estudio	• Probabilidad muy remota
• Más fácil de mantener al entrevistador ciego al grupo al que pertenece cada individuo	• Difícil de mantener el ciego

### Número de controles por caso

Cuando existe un número de casos suficiente, se suele seleccionar un control por cada uno. Cuando el número de casos es limitado se puede aumentar la potencia estadística del estudio para detectar un efecto determinado y seleccionar más de un control por cada caso. Esta última alternativa es muy útil cuando el coste de seleccionar a un control es relativamente bajo.

Como norma general, cuando la razón de controles por caso excede el valor de 4:1, la ganancia en términos de potencia estadística es muy pequeña comparada con el coste de seleccionar a los controles.

### Número de grupos control

Para valorar la posibilidad de que se haya cometido un sesgo en la selección de los controles, algunos autores recomiendan, si es posible, utilizar dos grupos control. Si se obtiene la misma estimación al comparar con cada uno de ellos, se podría asumir que no ha existido un sesgo de selección. A pesar de todo, siempre existe la posibilidad remota de que las dos estimaciones tengan el mismo grado de sesgo.

**Ejemplo 8.7.** El objetivo de un estudio era valorar el riesgo de padecer cáncer de vejiga asociado a la ocupación laboral, al consumo activo y pasivo de tabaco, la dieta, el consumo de café, analgésicos, edulcorantes y antecedentes de infección y litiasis urinaria (González et al, 1989). Para ello se estudiaron 497 casos y 583 controles hospitalarios, así como 530 controles de población. El porcentaje de encuestas no realizadas en los controles de población fue del 34,3%, ligeramente superior que en los controles hospitalarios (29,5%). Las causas por las que no se realizó la encuesta fueron, principalmente, la

dificultad para localizar los controles (mayor en el grupo poblacional) y la negativa a ser entrevistado (también mayor en el grupo poblacional). Para los controles hospitalarios utilizaron una lista amplia de diagnósticos de exclusión por ser enfermedades asociadas con los factores de estudio. Con el objeto de evitar que los controles pudieran concentrarse en alguna enfermedad asociada a las exposiciones, seleccionaron controles de la mayoría de servicios y con una amplia variedad de diagnósticos, como hernia inguinal, apendicitis, fractura de tibia, glaucoma, dolor dorsal y osteoartrosis, entre otros. Además, ninguno de estos diagnósticos superaba el 7% del total de los controles hospitalarios. Los autores defendían la inclusión de un control poblacional, además del hospitalario, como precaución frente a la eventualidad de que ciertos hábitos relacionados con los factores de estudio fueran diferentes en ambos grupos. Contar con dos grupos control tenía, además, el objetivo de aumentar la precisión y el poder estadístico del estudio. Al comparar ambos grupos por las variables consideradas de interés no se observaron diferencias significativas, lo que sugiere que ambos grupos proceden de una misma población.

La recomendación de usar dos grupos control es difícil de llevar a la práctica por su coste en tiempo y dinero. Además, algunos autores argumentan que los investigadores no suelen depositar igual confianza en la validez de las comparaciones basadas en dos o más grupos diferentes, y que con grupos de comparación distintos hay que esperar a que se observen resultados distintos. ¿Qué sucede si en un estudio largo y costoso se observan resultados discrepantes con distintos grupos de comparación? Si los investigadores depositan mayor confianza en uno de ellos, los resultados del otro restarían credibilidad a sus

hallazgos. En consecuencia, algunos epidemiólogos recomiendan escoger sólo un grupo de comparación, en general en el que se tenga mayor confianza.

## SESGOS EN LA SELECCIÓN DE CASOS Y CONTROLES

Un sesgo de selección aparece cuando los casos o los controles son incluidos o excluidos de un estudio debido a alguna característica relacionada con la exposición. Aunque los sesgos se explican con detalle en el capítulo correspondiente, a continuación se mencionan los más importantes.

El grupo de casos debe incluir todos los casos que aparecen en una población definida o bien una muestra representativa. Si se dispone de un registro de enfermedad de base poblacional, como los que suelen existir para cáncer o enfermedades congénitas, ésta será la fuente de elección, ya que se tendrá una lista exhaustiva de todos (o casi todos) los casos. Si no se dispone de este tipo de registros, debe recurrirse a otras fuentes, en general servicios hospitalarios u otros centros sanitarios. En esta situación, la inclusión de los casos está relacionada con la probabilidad de que hayan sido diagnosticados y admitidos en dichos servicios o centros. Si esta probabilidad depende de factores como el grado de especialización o el prestigio del servicio, o bien de características del paciente (lugar de residencia, clase socioeconómica, etc.), y estos factores están relacionados con el hecho de haber sido expuestos, la muestra de casos no será representativa y se obtendrá una estimación sesgada del efecto.

Otro sesgo de selección que puede presentarse en los estudios de casos y controles es el *sesgo de Berkson*. Suele ocurrir cuando la combinación de la exposición y la enfermedad en estudio aumenta la probabilidad de ingreso en un hospital, lo cual conduce a una frecuencia de exposición sistemáticamente más elevada en los casos que en los controles hospitalarios. El resultado es una estimación sesgada de la OR.

El uso de casos prevalentes en lugar de incidentes puede dar lugar a un error sistemático que se conoce con el nombre de *sesgo de Neyman o de supervivencia selectiva*. La prevalencia depende tanto de la duración de la enfermedad, que se verá afectada por el tratamiento y la atención sanitaria recibida, como de la letalidad de la enfermedad. Por estas razones, los casos prevalentes pueden no ser representativos de todos los casos.

El llamado *sesgo de detección* aparece como consecuencia de la existencia de una diferente probabilidad de diagnosticar a los casos y a los controles. Este sesgo mereció la atención de muchos investigadores a raíz

de la controversia surgida con el uso de estrógenos y el riesgo de padecer cáncer de endometrio. Se argumentó que la exposición (uso de estrógenos) estaba asociada a una mayor probabilidad de que los médicos detectaran la enfermedad, ya que las mujeres que utilizaban estrógenos eran visitadas con mayor frecuencia y de forma más cuidadosa que las que no los consumían. Sin embargo, la asociación entre el uso de estrógenos y el cáncer de endometrio permaneció después de considerar el posible efecto de un sesgo de detección.

## EMparejamiento (MATCHING)

El emparejamiento (*matching*) se refiere al proceso de seleccionar uno o más controles que se asemejen a un caso en los valores de determinadas variables, con la finalidad de controlar potenciales factores de confusión (p. ej., seleccionar para cada caso un control del mismo sexo y de edad similar).

El emparejamiento indirecto se utiliza cuando los potenciales factores de confusión son identificables pero difícilmente medibles. Si se quieren controlar determinadas variables socioeconómicas o ambientales se puede decidir equiparar a los casos y controles por el barrio de residencia, asumiendo que es una buena aproximación de aquellas variables.

Para que el emparejamiento sea útil, debe realizarse por una variable que sea un potencial factor de confusión, es decir, relacionada tanto con el factor de estudio (exposición) como con la variable de respuesta (enfermedad). En caso de que no sea así, se pierde eficiencia y se disminuye la validez de la comparación entre los casos y los controles, fenómeno que se conoce como sobreemparejamiento (*overmatching*). Este fenómeno también se produce cuando el emparejamiento no se acompaña de una técnica de análisis para datos apareados.

Aunque es una técnica que se ha usado ampliamente, sus múltiples inconvenientes, junto con el desarrollo de las técnicas de análisis multivariante, han disminuido su atractivo. Las ventajas y desventajas del emparejamiento se resumen en el cuadro 8.1.

Existen algunas situaciones en las que el emparejamiento puede ser útil e incluso deseable. Para determinadas variables, si durante el diseño del estudio no se emplea el emparejamiento, el control de su posible efecto de confusión se hace muy difícil. Por ejemplo, cuando se selecciona un caso según el barrio de residencia, que está correlacionado con multitud de variables (p. ej., estatus socioeconómico, hábitos dietéticos, utilización de los servicios sanitarios, etc.), es útil emparejarlo a un control que viva en el mismo lugar. Si los controles fueran seleccionados al azar de la población general y posteriormente se estudiara la

### Cuadro 8.1 Ventajas y desventajas del emparejamiento

#### Ventajas

- Es un proceso fácil de comprender
- Equilibra el número de casos y controles en cada categoría de la variable por la que se empareja
- Facilita la detección de una interacción entre la exposición y el factor por el que se aparea siempre que éste tenga un gran efecto sobre el riesgo de padecer la enfermedad y sea poco frecuente en la población
- Si el emparejamiento es perfecto, como sucede con las variables dicotómicas o las cuantitativas cuando se aparea exactamente por el mismo valor de la variable, el efecto de confusión se controlará completamente

#### Desventajas

- No se podrá estudiar el efecto de la variable por la que se empareja sobre la enfermedad
- Sólo se previene el posible efecto de confusión de las variables por las que se ha emparejado e, indirectamente, por aquellos factores que están muy relacionados con ellas
- Permite evaluar la existencia de interacción, pero no su estudio detallado
- Es un proceso laborioso que requiere mucho tiempo y puede alargar la duración del estudio
- Reduce la flexibilidad del análisis, al ser necesario utilizar técnicas para datos apareados
- Si se empareja por una variable que no es un factor de confusión se obtendrá una estimación poco precisa
- Aumenta la complejidad del estudio, a menudo con escaso beneficio, en términos de precisión y exactitud del estimador

asociación entre el barrio de residencia y la enfermedad, es muy probable que muy pocos controles procedieran del mismo barrio que los casos, por lo que el efecto de esta posible variable de confusión no quedaría controlado, por la imposibilidad de realizar un análisis estratificado o multivariante.

También es útil cuando el número de casos es muy pequeño, ya que en esta situación es muy posible que las características basales entre casos y controles difieran sustancialmente debido a variabilidad en el muestreo. Además, si la muestra es pequeña no permitirá un análisis de subgrupos, en el que además se tenga en cuenta el efecto de los posibles factores de confusión. El emparejamiento de varios controles por caso permitirá solucionar, al menos parcialmente, este problema.

También es útil cuando se desea controlar el efecto de un factor de confusión que se distribuye de forma muy desigual entre los casos y los controles.

**Ejemplo 8.8.** Supongamos un estudio que desea estimar el riesgo de infarto agudo de miocardio (IAM) asociado al consumo de tabaco en hombres de 20 a 44 años de edad. Dado que la edad está relacionada con la exposición (consumo de tabaco) y también con la enfermedad (IAM), puede considerarse un potencial factor de confusión. Si se opta por un diseño no emparejado y se selecciona una muestra aleatoria de controles, una forma de controlar el efecto de la edad es realizar un análisis estratificado. Otra forma de diseñar el estudio sería seleccionar una muestra de casos y emparejar los controles en distintas categorías de edad (p. ej.,  $\pm 2$  años). En el estudio sin emparejamiento los casos predominarán en las edades mayores, mientras que los controles se distribuirán uniformemente entre todos los grupos de edad. De hecho, en los grupos de edad más jóvenes habrá muy pocos casos, si es que hay alguno. Si en un grupo no existe ningún caso, la información sobre los controles en este grupo no tendrá ninguna utilidad. Además, las estimaciones del riesgo en los estratos será más imprecisa (tendrá una variancia mayor) en las situaciones donde existe un desequilibrio en el número de casos y controles.

Sin embargo, la reducción de la variancia que se consigue con el emparejamiento es modesta. Se suele alcanzar una reducción de entre un 5 y un 15% de la variancia del estimador. Las mayores reducciones se consiguen cuando existe una asociación muy importante entre la enfermedad en estudio y el factor de confusión (riesgo relativo [RR] > 20).

La conclusión es que el emparejamiento debe utilizarse casi exclusivamente en estudios de casos y controles y en contadas ocasiones. Existen técnicas de análisis que permiten controlar el efecto de los factores de confusión de forma mucho más eficiente.

### INFORMACIÓN SOBRE LA EXPOSICIÓN

La secuencia temporal causa-efecto de los estudios de casos y controles obliga a la obtención de información sobre la historia de exposiciones pasadas, es decir, de forma retrospectiva. Por esta razón, las encuestas, ya sean personales, telefónicas o mediante cuestionarios autoadministrados, son el método más utilizado para recoger dicha información. También pueden emplearse medidas biológicas asumiendo su estabilidad a lo largo del tiempo.

En ocasiones, se recoge la información a partir de personas próximas (familiares, amigos, etc.) a los sujetos incluidos en el estudio cuando éstos no son capaces de proporcionarla. El uso de informadores indirectos es muy útil cuando la enfermedad tiene una elevada letalidad y en los que la serie de casos es muy pequeña, ya que la falta de información de los individuos que han muerto puede causar un sesgo importante. Deben emplearse tanto en los casos como en los controles y, siempre que sea posible, validar la información sobre la exposición, comparando la dada por una muestra de sujetos participantes en el estudio con la ofrecida por su correspondiente informador indirecto.

La elección de la fuente de información depende de la exposición que se quiere medir. Si, por ejemplo, el interés está en determinar el número de cigarrillos y el tiempo que hace que fuma, lo mejor será preguntárselo directamente al individuo, ya que esta información no suele constar en las historias clínicas. Además, si el consumo de cigarrillos está relacionado con la enfermedad en estudio, probablemente se registre mucho mejor esta variable en los casos que en los controles.

**Ejemplo 8.9.** Supongamos un estudio en que los casos son pacientes afectados de bronquitis crónica, mientras que los controles son sujetos que padecen artrosis. Es muy posible que en las historias de los casos conste con detalle su hábito tabáquico, pero en la de los controles sólo esté registrado si fuma o no. En este caso, el grado con que se identifica la exposición en uno y otro grupo es muy diferente, lo que conducirá a unos resultados sesgados.

En otras ocasiones es mejor obtener la información de las historias clínicas, en especial cuando se trata de los resultados de pruebas realizadas a los pacientes.

**Ejemplo 8.10.** En un estudio se describió la concordancia entre la información dada por los pacientes y sus médicos con respecto a la práctica de la tinción de Papanicolaou y su resultado (Walter et al, 1988). Los datos procedían de un estudio de casos y controles recogidos mediante entrevista personal. Posteriormente, se contactó con sus respectivos médicos de cabecera o del centro de planificación familiar con el fin de verificar la información. Al comparar los hallazgos histológicos encontraron que existía bastante concordancia si se consideraban sólo como «normal-anormal», pero cuando se solicitaba el diagnóstico exacto de los resultados anormales la concordancia disminuyó sensiblemente. Las displasias eran clasificadas a menudo por las pacientes como cáncer, y viceversa. Las diferencias pueden deberse a que el médico ha proporcionado una in-

formación deficiente o incompleta, o simplemente a que las pacientes no reconocen las diferencias entre una displasia y una neoplasia invasiva. Cuando los autores calcularon los resultados por separado para los casos y los controles, se observó que en los primeros los datos eran más concordantes que en los segundos. Los casos estaban quizás más motivados para responder a este tipo de preguntas debido a que padecían una enfermedad grave. Además, es muy posible que les hicieran preguntas similares mientras duraba el proceso diagnóstico, y de este modo tuvieran más tiempo de considerar sus respuestas.

### Sesgo de memoria

Los estudios de casos y controles se han criticado muy a menudo a causa de que la información sobre la exposición se recoge retrospectivamente y, de este modo, se facilita la posibilidad de incurrir en un sesgo de memoria. Este sesgo suele ocurrir en enfermedades graves y/o cuando suponen un fuerte trauma psicológico, como las malformaciones congénitas. En estas circunstancias es muy posible que los casos recuerden sus antecedentes personales con mucho más detalle que los controles, al estar más sensibilizados por la enfermedad y porque sus médicos les habrán preguntado con insistencia por ellos. La presencia y la magnitud de este sesgo varían según la exposición.

**Ejemplo 8.11.** En un estudio cuyo objetivo era evaluar su presencia en madres de niños con malformaciones congénitas, se obtuvo información mediante entrevista personal sobre ocho posibles exposiciones durante el embarazo, y se comparó con la registrada en la historia clínica del obstetra. Se consiguió la misma información, y de las mismas fuentes, de un grupo de madres de hijos sin malformaciones, que sirvió de control (Werler et al, 1989). La información sobre cinco de las ocho exposiciones fue igual de exacta en los casos que en los controles, pero la historia de infección durante el embarazo, de infertilidad tratada o no, y de la utilización de métodos anticonceptivos durante 2 semanas o más después del último período menstrual, fue mucho más exacta en los casos que en los controles.

Estos resultados indican que el sesgo de memoria puede estar presente en algunas exposiciones y no en otras. Este hecho no es sorprendente, y probablemente se deba a una mezcla de memoria, emoción e impresiones de las personas sobre los riesgos asociados a un determinado factor. Sin embargo, es muy difícil poder determinar a priori en qué exposiciones es más posible que exista un sesgo de memoria.

La probabilidad de que se produzca este sesgo aumenta cuando se recoge información sobre varios factores de forma simultánea. Puede intentarse prevenir seleccionando como controles pacientes con enfermedades similares a la de estudio. Por ejemplo, en el caso de estudiar un determinado tipo de cáncer, el grupo control estaría formado por pacientes con otros cánceres, o si la enfermedad de interés es una malformación determinada, estaría constituido por madres de niños con otras malformaciones. Si se adopta esta estrategia, hay que tener la seguridad de que la exposición en estudio no está relacionada con las enfermedades que componen el grupo control, porque de otro modo se incuraría en un sesgo de selección.

### Sesgo del entrevistador

Puede aparecer siempre que el encuestador (u observador, en general) tenga conocimiento del grupo al que pertenece el sujeto al que está entrevistando y del objetivo del estudio. A la mayoría de los investigadores les gusta obtener resultados positivos y, de forma involuntaria, pueden preguntar con más detalle e insistencia a los casos que a los controles. Por ello, cuando la información se recoge a través de un cuestionario y siempre que los recursos lo permitan, es preferible que el encuestador sea alguien ajeno al equipo que ha diseñado el protocolo.

La forma de evitar este problema es que el encuestador no sepa si entrevista a un caso o a un control. Sin embargo, esto puede ser difícil de conseguir cuando los controles son personas sanas. Además, es muy posible que los casos antes o durante la entrevista digan alguna frase o comentario que haga sospechar al encuestador. De todos modos, siempre que sea posible se intentará que los entrevistadores permanezcan «ciegos» al grupo al que pertenece cada persona, e igualmente que no conozcan la hipótesis que se está evaluando.

## ESTUDIOS DE CASOS Y CONTROLES QUE EVALÚAN MEDIDAS PREVENTIVAS

Existen medidas preventivas muy utilizadas y aceptadas tanto por la población como por los profesionales sanitarios que nunca se han evaluado mediante un ensayo clínico. Por ejemplo, aunque se han publicado multitud de estudios que sugieren la efectividad del cribado para reducir la mortalidad por cáncer de cuello uterino, la mayoría de los resultados se sostienen en la comparación de las tasas entre poblaciones con diferente intensidad en sus programas de cribado o de una misma población durante distintos perío-

dos de tiempo. Dada la probabilidad de sesgo en los resultados de estos estudios, se podría dudar de que la relación entre la aplicación de un programa de despistaje y la reducción de la mortalidad por cáncer de cuello fuera causal. Aunque la evidencia debería proporcionarla un ensayo clínico, la realización de este estudio es difícil, ya que el cribado está considerado, tanto por el público general como por los profesionales sanitarios, una prueba útil.

Cuando la práctica de un método de cribado es muy frecuente, se puede utilizar el diseño de casos y controles para evaluar su eficacia. Si la medida es eficaz para prevenir una enfermedad y se ha aplicado en una población concreta, los individuos que la componen y que mueran a causa de esta enfermedad deberían haberse sometido a cribado con menos frecuencia que el resto de la población. Así pues, la comparación entre la historia de cribado de los casos que mueren por la enfermedad y la de los controles da una estimación del efecto protector del cribado.

Los estudios de casos y controles que evalúan la eficacia de un método de cribado tienen características propias. En primer lugar, el cribado está implicado directamente en el propio proceso, ya que conduce al diagnóstico precoz con independencia de que el tratamiento aplicado tempranamente sea o no eficaz. En consecuencia, la población sometida con regularidad a algún tipo de técnica de detección precoz tendrá una alta frecuencia de casos descubiertos durante el intervalo de detección (periodo de tiempo transcurrido entre la detección de la enfermedad por cribado y el momento en que habría sido diagnosticada por los cuidados médicos habituales en ausencia del cribado). Si una persona se considera caso en función de una característica de la enfermedad que se manifiesta durante ese intervalo, habrá un exceso de casos detectados por cribado y, por consiguiente, se enmascarará su posible efecto beneficioso.

Es necesario que los casos se elijan en función de las manifestaciones que desarrollen después del intervalo de detección. Una variable de respuesta que cumple esta condición es la mortalidad. El grupo de casos lo formarían aquellos individuos que murieron de la enfermedad y que se diagnosticaron como resultado del despistaje, así como los que murieron y cuya enfermedad se descubrió cuando ya había dado síntomas. En el grupo control se incluirían todos los miembros vivos (personas que no han padecido nunca la enfermedad más aquellos que sí la padecen) de la población de donde proceden los casos. A simple vista puede parecer extraño incluir a individuos con la enfermedad en el grupo control, pero si el diagnóstico temprano es eficaz, lo más probable es que las personas que sepan que padecen la enfermedad hayan sido sometidas a cribado. Si estos individuos se excluyeran de forma

sistemática, se tendería a reducir la estimación del efecto protector del cribado.

En ocasiones no resulta sencillo recoger la historia de exposición de los registros de los casos que han muerto. En esta situación se pueden elegir como casos a los individuos que se encuentran en estadios avanzados de la enfermedad. El grupo control estará formado por todas aquellas personas de la población de donde procedan los casos pero que no tengan la característica que los define, es decir, que no padeczan la enfermedad en estadios avanzados. Ello significa que entre los individuos potencialmente elegibles habrá personas sin la enfermedad, con independencia de si se les han aplicado o no medidas preventivas, y otras en las que la enfermedad estará menos avanzada, se haya o no descubierto después de un cribado.

Sin embargo, un error frecuente es elegir como grupo control sólo a pacientes con manifestaciones poco avanzadas de la enfermedad y comparar su historia de exposición con la del grupo de casos. Esta comparación no es apropiada, ya que con toda probabilidad su historia de cribado no será representativa de la población de donde provienen los casos. En la mayoría de las ocasiones la enfermedad se habrá detectado gracias al cribado, por lo que su exposición será superior, en promedio, a la de la población.

## **VENTAJAS Y LIMITACIONES DE LOS ESTUDIOS DE CASOS Y CONTROLES**

La mayor ventaja de los estudios de casos y controles es el grado de información que ofrecen, ya que se pueden estudiar un gran número de casos de una enfermedad, mientras que en un estudio de cohortes sólo unos pocos la desarrollarán. Esta ventaja se ve acrecentada cuando se estudian enfermedades poco frecuentes. Además, permiten evaluar varios factores de riesgo de forma simultánea y la existencia de interacciones entre ellos.

Una segunda ventaja es su eficiencia, ya que se realizan en mucho menos tiempo y con un coste menor que los estudios de cohortes, al no tener que realizar el seguimiento de los pacientes durante el período de latencia de la enfermedad.

Por último, permiten estudiar exposiciones que son raras en la población general, siempre y cuando estén asociadas a la enfermedad en estudio. Por ejemplo, la exposición a las anilinas es rara en la población general, pero está relacionada con el cáncer de vejiga, por lo que es una exposición que podrá investigarse de manera eficiente mediante un estudio de casos y controles.

Su limitación más importante es que existe un riesgo elevado de introducción de sesgos tanto en la selección de los grupos como en la información que se obtiene sobre los factores de riesgo. Los estudios de cohortes, al elegir la población a partir de la exposición, están menos sujetos a la introducción de errores sistemáticos.

Los estudios de casos y controles no proporcionan una estimación directa de la incidencia ni de la prevalencia de una enfermedad, ya que la proporción de los participantes en el estudio que tienen la exposición está determinada por el investigador, no por la proporción que existe en la comunidad. Una excepción son los estudios poblacionales en los que se asume que la tasa del grupo control es representativa de toda la población y además se conoce su incidencia. A partir de esta información puede calcularse la incidencia en el grupo expuesto y no expuesto (cuadro 8.2). Los estudios de casos y controles permiten estimar directamente el riesgo de padecer una enfermedad asociado a un determinado factor de riesgo.

El cuadro 8.3 resume las ventajas y limitaciones de estos estudios. En general, en la exploración inicial de una hipótesis son mucho más útiles los estudios de casos y controles por su eficiencia en tiempo, coste y posibilidad de estudiar varios factores de forma simultánea. En el caso de que se requiera una mayor evidencia, los estudios de cohortes, siempre que sean factibles, ofrecen datos sobre la evaluación de una asociación de una manera más directa y con una metodología distinta a la aplicada a los estudios de casos y controles.

## **DISEÑOS HÍBRIDOS**

Los diseños híbridos tienen características tanto de los estudios de cohortes como de los de casos y controles, pero obvian algunas de sus desventajas. Se caracterizan por analizar todos los casos aparecidos en una cohorte estable seguida en el tiempo y utilizar como controles sólo una muestra de los sujetos de esa misma cohorte. Según el plan de muestreo que se utilice para constituir los grupos a partir de los componentes de la cohorte, pueden diferenciarse dos tipos generales de diseños: estudios de casos y controles anidados dentro de una cohorte y estudios de cohorte y casos.

Dado que la incidencia de la mayor parte de las enfermedades que se estudian es relativamente baja, interesa seleccionar todos los casos que aparecen en la cohorte, aunque podría utilizarse cualquier otra fracción de muestreo. Por el contrario, los controles serán una muestra de sujetos procedentes de la cohorte inicial, con lo que se pierde precisión estadística; sin embargo, este hecho se compensa por la considerable

### Cuadro 8.2 Cálculo de las tasas de incidencia a partir de un estudio de casos y controles

Para el cálculo de las tasas de incidencia es imprescindible conocer la incidencia de la enfermedad en toda la población (p. ej., a partir de registros de enfermedad) y asumir que la prevalencia de exposición en el grupo control es representativa de toda la población.

La incidencia en toda la población ( $It$ ) es igual a la media ponderada de las incidencias según varios grados de exposición, y el factor de ponderación es la proporción de individuos en cada grado. De este modo, la  $It$  es la suma de la incidencia en el grupo expuesto ( $Ie$ ) multiplicado por la proporción de expuestos en toda la población ( $Pe$ ) y de la incidencia en el grupo no expuesto ( $Io$ ) multiplicado por la proporción de no expuestos en la población ( $Po$ )

$$It = (Ie \times Pe) + (Io \times Po)$$

Dado que el riesgo relativo (RR) es la razón de incidencias entre el grupo expuesto y el no expuesto, la incidencia en el grupo expuesto será igual a:

$$Ie = RR \times Io$$

Por tanto:

$$It = (RR \times Io \times Pe) + (Io \times Po)$$

En un estudio de casos y controles, la RR se puede sustituir por la *odds ratio* (OR):

$$It = (OR \times Io \times Pe) + (Io \times Po)$$

Despejando, se obtendrá la incidencia en el grupo no expuesto ( $Io$ )

$$Io = \frac{It}{(OR \times Pe) + Po}$$

Una vez calculada la incidencia en el grupo no expuesto, basta multiplicarla por la OR y se obtendrá la incidencia en el grupo expuesto

reducción en el número de sujetos estudiados y del coste en la recogida de datos y el tiempo de seguimiento, y por un mayor control de calidad de la recogida y el procesamiento de los datos.

### Estudios de casos y controles anidados en una cohorte

A partir de un estudio de cohortes ya realizado o que se está llevando a cabo se identifican todos los sujetos que han desarrollado la enfermedad, que constituirán el grupo de casos. En el momento en que aparece un caso se selecciona aleatoriamente uno o varios controles entre los sujetos de riesgo en ese momento

### Cuadro 8.3 Ventajas y desventajas de los estudios de casos y controles

#### Ventajas

- Útiles para estudiar enfermedades poco frecuentes
- Son un buen diseño para estudiar enfermedades con largos períodos de latencia
- Permiten estudiar exposiciones poco frecuentes siempre que estén asociadas a la enfermedad
- Pueden evaluar múltiples factores de riesgo para una sola enfermedad
- Su duración es relativamente corta

#### Desventajas

- Es fácil que se introduzcan errores sistemáticos tanto en la selección de los grupos como al recoger la información
- En ocasiones es difícil establecer la secuencia temporal entre la exposición y la enfermedad
- No son un buen diseño para estudiar más de una enfermedad de forma simultánea
- No permiten calcular la incidencia o prevalencia de la enfermedad

(plan de muestreo por densidad de incidencia). Los controles pueden emparejarse con los casos, y es conveniente hacerlo por alguna variable dependiente del tiempo, como los años de permanencia en la cohorte. Además, un mismo sujeto podría seleccionarse como control en más de una ocasión para casos diferentes, o podría seleccionarse como control en un momento determinado y considerarse como caso posteriormente si desarrolla la enfermedad.

Este diseño está indicado cuando es necesario realizar mediciones muy costosas. Comparten con los estudios de cohortes el hecho de utilizar como estudio a un grupo definido de personas (cohorte) que son seguidas durante un período de tiempo, y el hecho de que la información sobre los factores de riesgo de interés y las variables principales se han recogido al inicio del seguimiento, de forma prospectiva y antes de que se desarrolle la enfermedad, lo que elimina algunos sesgos de información que hubieran podido aparecer si se hubiera utilizado un diseño clásico de casos y controles. Además, comparten con los de casos y controles el plan de muestreo: se recogen todos los casos de la enfermedad y una muestra de controles representativa de la población de la que proceden los casos con riesgo de enfermedad, pero que en el momento en que se seleccionaron como controles aún no la habían desarrollado.

**Ejemplo 8.12.** Un ejemplo de un estudio de casos y controles anidado es un estudio que investigaba la relación entre las concentraciones séricas de colesterol y el cáncer de colon (Sidney et al, 1986). La cohorte en que se anidaba el estudio consistía en 48.314 individuos a los que se había practicado un examen de salud, y que fueron seguidos durante un período promedio de 7,2 años (más de 348.000 personas-año) para observar el desarrollo de cáncer de colon. Los 245 individuos de la cohorte que desarrollaron la enfermedad se consideraron casos, y para cada uno de ellos se seleccionaron cinco controles. Los casos y los controles se emparejaron por la edad, el sexo y el tiempo transcurrido desde el primer examen. En lugar de tener que analizar el colesterol y otras variables en todos los miembros de la cohorte, los investigadores concentraron todos sus esfuerzos en los 245 casos y sus controles, una muestra mucho más pequeña y asequible.

### Estudios de cohorte y casos

En estos estudios se selecciona aleatoriamente una muestra (subcohorte) de la cohorte inicial, que servirá como grupo de comparación para todos los casos que aparezcan durante el seguimiento, con independencia de si ya pertenecían o no a la subcohorte. En resumen, se eligen todos los casos de la cohorte inicial aparecidos durante el seguimiento, y su información se compara con la procedente de una muestra de la cohorte inicial. Lo que se persigue es obtener una nueva cohorte, con menos sujetos que la inicial, en la que los casos están sobrerepresentados. Esta sobrerepresentación de los casos debe tenerse en cuenta en el análisis, ya que no podrá analizarse como si se tratara de un estudio de cohorte tradicional.

La misma subcohorte puede servir de grupo de comparación para el estudio de diversas enfermedades. Si

se deseara estudiar la relación entre determinados factores de riesgo y la aparición de dos enfermedades bastaría con comparar los casos de cada enfermedad con la misma subcohorte seleccionada.

Este diseño permite determinar las tasas de incidencia de la enfermedad y no sólo el riesgo relativo.

**Ejemplo 8.13.** Un ejemplo de diseño de cohorte y casos es un estudio que estimaba la relación entre el consumo de alcohol y el desarrollo de cáncer de mama (Van der Brandt et al, 1995). La cohorte inicial estaba constituida por 62.573 mujeres de entre 55 y 69 años de edad en las que se evaluaban aspectos relacionados con la dieta y el cáncer. Los investigadores seleccionaron de forma aleatoria una subcohorte de 1.812 mujeres, que fueron seguidas bianualmente para recoger información detallada sobre covariables de interés. Los casos fueron las 422 mujeres que desarrollaron cáncer de mama durante el seguimiento de la cohorte inicial. Estos casos podían proceder de la subcohorte o de la cohorte inicial. Nótese que se ha podido recoger información más exacta y detallada de 1.812 mujeres que de las 62.573 de la cohorte inicial. Además, estas 1.812 mujeres pueden servir de denominador no sólo para evaluar hipótesis relacionadas con el cáncer de mama sino con otros tipos de cáncer.

La elección de una u otra alternativa de diseño depende sobre todo del tipo de datos que el investigador necesita procesar. Si el procesamiento o ciertas características de los datos cambian con el tiempo, son más ventajosos los diseños anidados. Por el contrario, si se puede asumir la estabilidad de la información en el tiempo, la posibilidad de estudiar múltiples enfermedades con el mismo grupo de comparación hace que los diseños de cohorte y casos sean de elección.

### BIBLIOGRAFÍA DE LOS EJEMPLOS

- |  |   |  |
|--|---|--|
| Colditz GA, Bonita R, Stampfer MJ, Willett WC, Rosner B, Speizer FE, et al. Cigarette smoking and risk of stroke in middle aged women. <i>N Engl J Med.</i> 1988;318:937-41. | cáncer de vejiga en España. <i>Med Clin (Barc).</i> 1989;92:646-51.   | Walter SD, Clarke EA, Hatcher J, Stitt LW. A comparison of physician and patient of reports of pap smears histories. <i>J Clin Epidemiol.</i> 1988;4:401-10. |
| González CA, López-Abente G, Errézola M, Escolar A, Izazugaza I, Riboli E, et al. Diseño y realización de un estudio multicéntrico caso control sobre                        | Sidney S, Friedman GD, Hiatt RA. Serum cholesterol and large bowel cancer. <i>Am J Epidemiol.</i> 1986;124:33-8.  | Werler M, Pober B, Nelson K, Holmes L. Reporting accuracy among mothers of malformed and nonmalformed infants. <i>Am J Epidemiol.</i> 1989;129:415-21.       |
|  | Van der Brandt PA, Goldbohm RA, Van't Veer P. Alcohol and breast cancer: results from the Netherlands Cohort Study. <i>Am J Epidemiol.</i> 1995;141:907-15. |  |

**BIBLIOGRAFÍA**

- Bland JM, Altman DG. The odds ratio. *BMJ*. 2000;320:1468.
- Caro JJ, Huybrechts KF. Case-control studies in pharmacoeconomic research: an overview. *Pharmacoeconomics*. 2009;27(8):627-34.
- Delgado-Rodríguez M. Discordancias entre los estudios de ámbitos hospitalario y comunitario cuando evalúan la misma pregunta de investigación. *Gac Sanit*. 2002;16:344-53.
- DiPietro NA. Methods in epidemiology: observational study designs. *Pharmacotherapy*. 2010;30:973-84.
- Fernández E. Estudios observacionales. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo científico. Barcelona: Elsevier; 2010. p. 205-14.
- Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet*. 2002;359:248-52.
- Langholz B, Richardson D. Are nested case-control studies biased? *Epidemiology*. 2009;20(3):321-9.
- Manchikanti L, Singh V, Smith HS, Hirsch JA. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: part 4: observational studies. *Pain Physician*. 2009;12(1):73-108.
- Maclure M, Schneeweiss S. Causation of bias: the episcope. *Epidemiology*. 2001;12: 114-22.
- Schlesselman JJ. Case-control studies: design, conduct, analysis. New York: Oxford University Press; 1982.
- Schulz KF, Grimes D. Case-control studies: research in reverse. *Lancet*. 2002;359:431-4.
- Silverman SL. From randomized controlled trials to observational studies. *Am J Med*. 2009;122(2):114-20.
- Sturmer T, Brenner H. Degree of matching and gain in power and efficiency in case-control studies. *Epidemiology*. 2001;12:101-8.
- Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. II. Types of controls. *Am J Epidemiol*. 1992;135:1029-41.
- Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. III. Design options. *Am J Epidemiol*. 1992;135:1042-50.
- Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med*. 2007;4(10):e297.
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453-7.

# Capítulo

9

## Estudios descriptivos

Las principales finalidades de los estudios descriptivos son describir la frecuencia y las características de un problema de salud en una población, describir la asociación entre dos o más variables sin asumir una relación causal entre ellas y generar hipótesis razonables que deberán contrastarse posteriormente mediante estudios analíticos.

Los estudios descriptivos son observacionales, y pueden ser tanto transversales como longitudinales, clasificándose según su objetivo. A continuación se presentan las características principales de los más habituales.

### ESTUDIOS DE PREVALENCIA Y DE ASOCIACIÓN CRUZADA

Los estudios de prevalencia y de asociación cruzada son diseños descriptivos transversales que tienen como finalidad principal la estimación de la prevalencia (anexo 1) de una enfermedad o una característica en una población. Se basan en la definición precisa de una población de estudio y la obtención de una muestra representativa de ella, en la que se determina la frecuencia de individuos que presentan la característica o enfermedad de estudio. La validez de los resultados dependerá de la representatividad de la muestra, de la calidad de los datos obtenidos y de que el número de personas en las que no se haya podido determinar la presencia de la enfermedad («no respuestas») sea pequeño y sus motivos no estén relacionados con ella.

En ocasiones, los diseños transversales se utilizan para estudiar la relación entre dos o más variables en una población en un momento del tiempo (*estudios de asociación cruzada*). Dado que las variables

se miden de forma simultánea o en un corto período, existe una ambigüedad temporal que dificulta en muchas ocasiones la interpretación de una posible relación causa-efecto. Además, el valor de una variable en un individuo en el momento en que se realiza el estudio puede ser muy diferente del que tenía un tiempo atrás. Esta limitación es menos importante cuando se estudian características invariables, como el grupo sanguíneo o el sexo, o factores en los que existe una buena correlación entre los valores actuales y los previos, como las preferencias dietéticas.

**Ejemplo 9.1.** Un estudio estima la prevalencia de artrosis de la rodilla en una comunidad y, a su vez, la relaciona con la presencia de obesidad (Davis et al, 1988). Para ello, se extrae una muestra representativa y se estudia si una persona tiene artrosis de rodilla, según unos criterios establecidos previamente, y a su vez, si es obesa o no. A continuación se compara la proporción de personas con artrosis que son obesas con la de las personas sin artrosis que son obesas. Si se encuentra una asociación positiva, como sucedió en el estudio que nos ocupa, existen varias hipótesis para explicarla. La teoría más aceptada es que la obesidad comporta una sobrecarga mecánica sobre la articulación, y acelera de este modo el deterioro del cartílago. Por otro lado, hay autores que piensan que la obesidad es una consecuencia, más que la causa de la artrosis, ya que ésta, al producir una limitación de la movilidad y el ejercicio, probablemente conduce a la obesidad. Como comentan los autores, los datos sólo sugieren que la obesidad es una causa, ya que con un estudio transversal, la hipótesis de que la obesidad antecede a la aparición de la artrosis no puede probarse, porque ambas variables se miden a la vez.

Cuando una de las variables es la presencia de una enfermedad, hay que tener en cuenta que se estudian casos prevalentes y que, por tanto, existe una mayor proporción de casos con enfermedad de larga duración (fig. 9.1). Si las características de los individuos cuya enfermedad es de corta duración o rápidamente mortal son diferentes de aquellas en que la enfermedad es de larga duración, puede existir un sesgo de supervivencia que conduce a estimaciones sesgadas de la relación entre algunas variables y la enfermedad (fig. 9.2).

**Ejemplo 9.2.** En un estudio se compararon las características clínicas de 1.135 pacientes recién diagnosticados de diabetes mellitus desde 1945 hasta 1969, con la de los casos prevalentes en el año 1970 en esta misma comunidad, con los siguientes resultados (Melton et al., 1983): los pacientes prevalentes tenían mayor edad y la proporción hombre:mujer era menor que la observada en los casos incidentes. Además, los casos prevalentes presentaban una mayor probabilidad de tomar hipoglucemiantes orales, tenían valores de glucosa en ayunas menores, menor probabilidad de estar sintomáticos, pero mayor probabilidad de padecer complicaciones macrovasculares y microvasculares. Estas diferencias

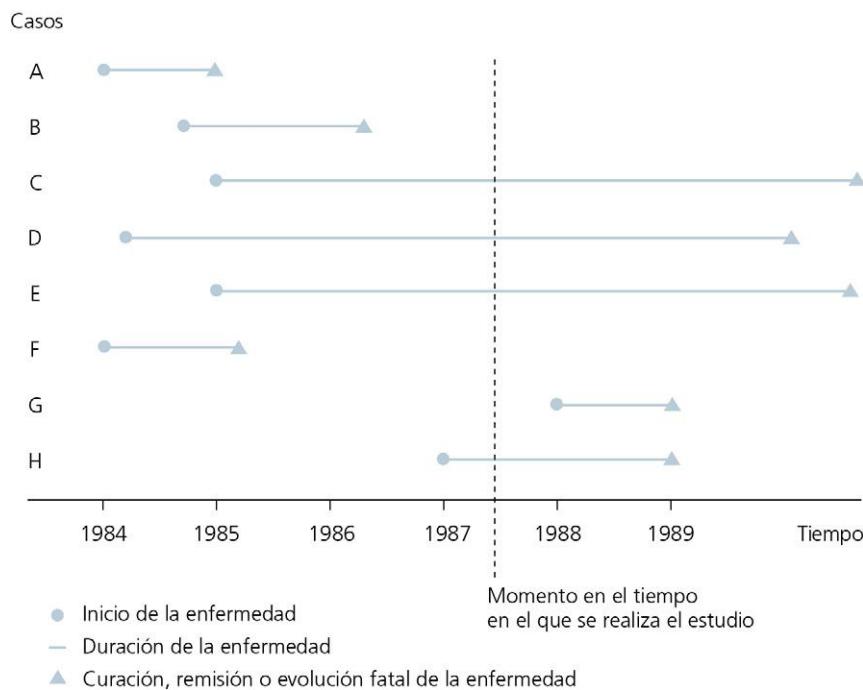
eran el resultado de la diferente tasa de supervivencia. Por todo ello, los resultados de los estudios de asociación cruzada se han de interpretar con suma cautela.

Las ventajas y desventajas de estos estudios se resumen en el cuadro 9.1.

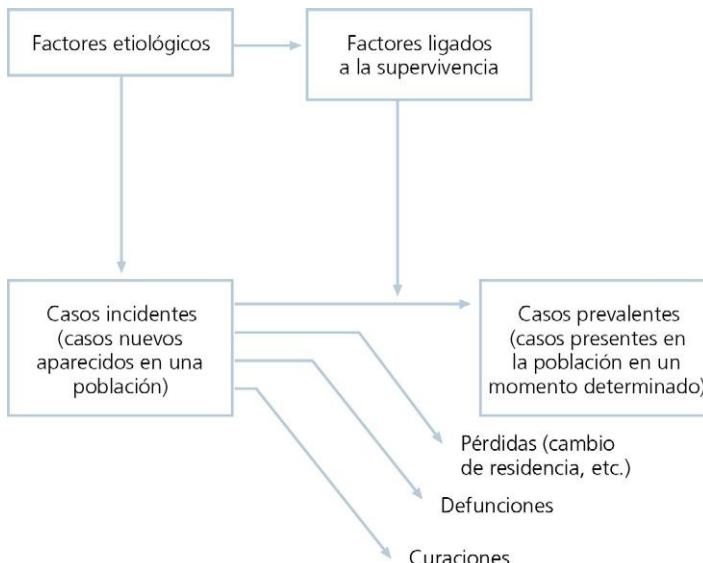
### Características de la enfermedad

La enfermedad estudiada debe ser de inicio lento y de larga duración, ya que, en caso contrario, existirán pocos individuos que la padecan en un momento determinado. Estos estudios no son útiles en el caso de enfermedades poco frecuentes, ya que para encontrar un número suficiente de casos debería recogerse información de una gran cantidad de personas.

Las variables estudiadas deben ser fácilmente medibles a través de cuestionarios o exámenes médicos simples y seguros, ya que se estudia una muestra de la población que en su mayoría no tendrá la enfermedad, por lo que no pueden usarse métodos que supongan algún riesgo para los participantes. Las pruebas deben ser lo más sensibles y específicas posible para evitar la clasificación incorrecta de sujetos en situación de remisión o en tratamiento.



**Figura 9.1** Estudio transversal y su relación con la duración de los casos.



**Figura 9.2** Diferencias entre casos prevalentes y casos incidentes.

#### Cuadro 9.1 Ventajas y desventajas de los estudios de prevalencia y asociación cruzada

##### Ventajas

- Suelen estudiar muestras representativas de la población de estudio, lo que facilita la extrapolación de resultados
- Se pueden estudiar varias enfermedades o factores de riesgo en un solo estudio
- Se realizan en un corto período de tiempo
- Permiten estimar la prevalencia y ayudan a la planificación de los servicios sanitarios
- Son un primer paso en la realización de muchos estudios prospectivos

##### Desventajas

- Falta de una secuencia temporal (dificultades para evaluar relaciones causa-efecto)
- No son útiles para enfermedades poco frecuentes
- Posibilidad de que exista un sesgo de supervivencia
- Posibilidad de sesgo debido a las no respuestas

**Ejemplo 9.3.** Si se desea conocer la prevalencia de la hipertensión arterial (HTA) en una población y ésta se define a partir de un determinado grado de presión arterial, es posible que muchos individuos ya

diagnosticados y tratados presenten cifras inferiores a las de referencia para definir la HTA. Si esto no se tiene en cuenta, se infraestimarán la verdadera prevalencia de la enfermedad. Por otro lado, la consideración de hipertensos por el mero hecho de estar en tratamiento también puede tener sus inconvenientes, ya que probablemente no han seguido los mismos criterios diagnósticos que se han establecido para la realización del estudio.

##### No respuestas

Si la prevalencia entre los que no participan difiere de la de los que sí lo hacen, se producirá un sesgo en la estimación del resultado.

**Ejemplo 9.4.** La tabla 9.1 recoge los resultados de un ejemplo hipotético. Se observa que la prevalencia estimada de hipercolesterolemia sería del 15% si sólo se tuvieran en cuenta los resultados de los participantes. Sin embargo, la prevalencia real en el total de la muestra es del 21%. Las diferencias aparecen porque el 30% de las no respuestas padecía la enfermedad.

No existe un porcentaje mínimo de no respuestas a partir del cual un estudio pueda considerarse no válido. La posibilidad de obtener resultados sesgados depende de la diferencia en la prevalencia entre los participantes y los no participantes, aunque, lógicamente, es mayor cuanto mayor es el porcentaje de no respuestas.

**Tabla 9.1 Resultados de un estudio hipotético sobre la prevalencia de la hipercolesterolemia (ejemplo 9.4)**

<b>Muestra de la población</b>	<b>Hipercolesterolemia</b>		
	<b>n</b>	<b>n</b>	<b>(%)</b>
Participantes	600	90	(15)
No participantes	400	120	(30)
Total	1.000	210	(21)

Deben utilizarse todas las estrategias posibles para reducir el porcentaje de no respuestas. Además, si se producen, es conveniente estudiar una muestra con el fin de averiguar si los sujetos que no han respondido tienen características distintas de los que sí lo han hecho y recoger la máxima información de los no participantes con el propósito de poder juzgar si los datos del estudio son válidos, e interpretarlos de forma correcta.

## SERIES DE CASOS

Las series de *casos transversales* consisten en la enumeración descriptiva de unas características seleccionadas, observadas en un momento del tiempo, en un grupo de pacientes con una enfermedad determinada o en un grupo de sujetos que tienen una determinada condición en común. Por ejemplo, describir las cifras actuales de presión arterial, colesterol y otros factores de riesgo cardiovascular en los diabéticos del centro de salud.

En las series de *casos longitudinales* o *seguimiento de una cohorte* se describe la evolución temporal de determinadas características observadas en un grupo de pacientes con una enfermedad o en un grupo de sujetos que tienen una determinada condición en común. Existe una secuencia temporal definida, pero el estudio no evalúa ninguna relación causa-efecto entre las variables.

## Seguimiento de una cohorte

Se trata de estudios descriptivos longitudinales que tienen como objetivo estudiar la evolución natural de una enfermedad para conocer la estabilidad o el cambio de ciertas características, describir los resultados observados en un grupo de pacientes que han recibido un tratamiento o intervención dentro del proceso habitual de atención y que no ha sido administrado de forma deliberada para la realización del estudio, o estimar la incidencia (anexo 1) de una enfermedad en un grupo de sujetos.

Estos estudios tienen dos características comunes. La primera es que la información sobre la enfermedad o la aparición de cualquier fenómeno se recoge con el tiempo, por lo que se puede relacionar, por ejemplo, con la duración del trastorno, los cambios en la gravedad, los tratamientos previos o las enfermedades concomitantes. La segunda es que describen la evolución de una muestra de sujetos que tienen una característica en común (p. ej., padecen la misma enfermedad). Es importante definir la cohorte e indicar qué sujetos se incluyen, de dónde proceden y el tiempo empleado para reclutarlos. A menudo se utilizan datos retrospectivos procedentes de registros clínicos, con las limitaciones que ello comporta.

Además de las no respuestas, en los estudios longitudinales se presenta el problema de las pérdidas y los abandonos durante el seguimiento. Si los motivos por los que se producen están relacionados con el problema estudiado, es decir, si la incidencia o el pronóstico es distinto entre los que abandonan y los que no lo hacen, se obtendrá una estimación sesgada del resultado.

## ESTUDIOS QUE EVALÚAN UNA PRUEBA DIAGNÓSTICA

El objetivo de estos estudios es estimar la capacidad de una medida (prueba diagnóstica) para discriminar entre las personas que padecen una enfermedad y aquellas que no la padecen, pero presentan unos síntomas similares. Se asume que el problema de salud puede medirse de forma válida y fiable mediante un procedimiento de referencia o *gold standard* y se desea evaluar la utilidad de una nueva prueba.

La estrategia de investigación consiste en administrar en una muestra de individuos tanto el procedimiento de referencia como el de estudio. El primero permitirá diferenciar los sujetos que padecen la enfermedad de los que no la padecen, de forma que se podrá evaluar la capacidad de la nueva prueba

diagnóstica para discriminar entre ambos grupos. Los resultados permiten clasificar a los sujetos estudiados en cuatro grupos en una tabla  $2 \times 2$ , a partir de la que se calculan la sensibilidad, la especificidad y otros índices de utilidad de la prueba (anexo 3).

Antes de iniciar el estudio debe evaluarse la reproducibilidad de la prueba, es decir, si produce los mismos resultados cuando se administra en más de una ocasión a pacientes estables. Una pobre reproducibilidad puede deberse a la propia naturaleza de la prueba, o a que se requiere habilidad en su realización o en la interpretación clínica de los resultados, sobre todo en aquellas pruebas cuyos resultados dependen de la subjetividad del observador. Una prueba poco reproducible puede afectar de forma grave a la validez de los resultados.

### Selección de los sujetos

Los pacientes en los que se evalúa una nueva prueba no han de diferir sustancialmente de la población a la que se aplicará en la práctica clínica. Si se estudiaran sólo aquellos que presentasen un proceso más grave, o más extenso, la nueva prueba daría más resultados positivos que si el grupo lo compusieran pacientes menos graves. Por esta razón, en el grupo de estudio se han de incluir pacientes con diferentes estados clínicos y patológicos de la enfermedad. Así mismo, es importante que no se excluyan aquellos sujetos que se encuentren en el límite diagnóstico.

El grupo de no enfermos permite determinar los falsos positivos. En este grupo se deben incluir individuos que tengan enfermedades que planteen problemas de diagnóstico diferencial con los pacientes del otro grupo. Si se incluyen sólo individuos sanos, existe el peligro de encontrar un número menor de falsos positivos, lo que se traduce en una especificidad de la prueba más elevada.

Existen diferentes estrategias para seleccionar la muestra o las muestras de individuos que se incluirán en el estudio.

### Selección de una sola muestra

Consiste en seleccionar una única muestra representativa de los sujetos a los que en la práctica se les aplicaría la prueba para realizar el diagnóstico de la enfermedad, y aplicar a todos ellos el procedimiento de referencia y la nueva prueba simultáneamente. Esta estrategia utiliza un diseño transversal muy similar al de los estudios de asociación cruzada.

**Ejemplo 9.5.** Un estudio tenía por objetivo evaluar qué variables clínicas y analíticas discriminaban a los pacientes que presentaban neumonía. Se seleccionaron

141 pacientes consecutivos que habían ingresado en urgencias en un hospital con fiebre de más de  $38^{\circ}\text{C}$  en las últimas 48 horas, tanto si tenían síntomas respiratorios como si no tenían ninguna focalidad. A estos 141 pacientes se les midieron las variables clínicas y analíticas de interés y, a continuación, se les aplicó el procedimiento de referencia para clasificarlos en enfermos (sujetos con neumonía) y no enfermos (sujetos sin neumonía).

Esta estrategia tiene ventajas de validez porque, una vez definidos los criterios de selección de la población de estudio, es de esperar que se encuentren individuos con un espectro amplio de la enfermedad, así como sujetos con signos y síntomas que impliquen un diagnóstico diferencial. Sin embargo, al tener que administrar a todos los pacientes la nueva prueba y la de referencia, el estudio puede resultar muy caro si el tamaño muestral requerido es muy amplio. Además, en determinadas ocasiones, la aplicación de la prueba de referencia puede suponer riesgos excesivos, lo que puede plantear problemas si debe aplicarse a un número elevado de sujetos.

### Selección de dos muestras a partir del diagnóstico

Se parte de dos grupos, uno formado por pacientes con la enfermedad diagnosticada a partir del criterio de referencia y otro formado por individuos sin ella. A todos ellos se les aplica la prueba en estudio y se comparan los resultados obtenidos. Este diseño es similar al de los estudios de casos y controles, en el sentido de que la enfermedad clasifica a los sujetos y la prueba se explora a continuación.

Como se ha comentado, es importante que en el grupo control se incluyan pacientes con situaciones que planteen problemas de diagnóstico diferencial con la enfermedad en estudio. Por ejemplo, si se desea evaluar una prueba diagnóstica del cáncer de colon, el grupo control debería incluir pacientes con enfermedades como la colitis ulcerosa o la enfermedad de Crohn, que pueden plantear problemas diagnósticos con el cáncer de colon, así como pacientes con otros tipos de cáncer.

**Ejemplo 9.6.** Cuando se evaluó originalmente el antígeno carcinoembionario (CEA, según sus siglas en inglés) en el diagnóstico precoz del cáncer de colon, se hizo en 36 pacientes con cáncer en grado avanzado y la prueba resultó positiva en 35 de ellos (Thomson et al, 1969). Al mismo tiempo, se observó que los valores de CEA eran mucho más bajos en sujetos asintomáticos. Los resultados sugerían que el CEA podía ser útil en el diagnóstico del cáncer

colorrectal. En estudios posteriores, realizados en pacientes con cáncer colorrectal en estadios no tan avanzados, comparando los resultados con los obtenidos en otros cánceres gastrointestinales, el CEA mostró una validez mucho menor y se dejó de utilizar como prueba de cribado (Bates, 1991).

Con esta estrategia, la prevalencia de la enfermedad en la población de estudio (la relación entre el grupo de enfermos y el de no enfermos) puede ser muy diferente de la que existe en la práctica. Así, por ejemplo, si se incluye el mismo número de enfermos que de no enfermos, la prevalencia en los sujetos estudiados es del 50%. Si bien permite la determinación de la sensibilidad, de la especificidad, de los cocientes de probabilidad y de las curvas ROC de forma eficiente, dificulta la estimación de los valores predictivos (anexo 3), ya que éstos dependen de la prevalencia de la enfermedad, y ésta depende a su vez del número de sujetos incluidos en cada grupo.

Respecto a la anterior, esta estrategia tiene una mejor relación coste-efectividad si el coste de las pruebas es elevado. Por tanto, es el diseño que se debe considerar cuando se trata de evaluar exploraciones de alta tecnología.

### Selección de dos muestras a partir del resultado de la prueba

Se parte de un grupo de individuos a los que se aplica la nueva prueba, y se obtienen dos subgrupos de sujetos: uno con resultado positivo y otro con resultado negativo. A continuación se aplica a cada subgrupo el procedimiento de referencia para comprobar la existencia de enfermedad. Esta estrategia se asemeja más a un estudio de cohortes. Tiene la ventaja de que los valores predictivos pueden calcularse directamente, pero dificulta la estimación de la sensibilidad y la especificidad.

### Selección del criterio de referencia

El criterio de referencia se utiliza para diferenciar entre enfermos y no enfermos y, por tanto, corresponde a la «verdad» contra la que se va a comparar la nueva prueba diagnóstica. Por ello, es muy importante definir con claridad la enfermedad o el problema de salud que se desea estudiar y elegir el criterio de referencia adecuado.

**Ejemplo 9.7.** Para estudiar la enfermedad coronaria pueden distinguirse diferentes conceptos: enfermedad coronaria aterosclerosa, que es un concepto anatómico; angina de pecho, que es un concepto clínico, o isquemia miocárdica, que es un concepto fisiopatológico.

Los tres están muy relacionados, aunque no son exactamente idénticos. Cada uno de ellos puede ser medido a través de un criterio de referencia distinto. Por ejemplo, para la enfermedad coronaria aterosclerosa, el criterio debería ser anatomopatológico; para la angina de pecho, sería un juicio clínico realizado preferiblemente por un panel de expertos tras una historia clínica, y para la isquemia miocárdica, el mejor método sería el resultado de las determinaciones metabólicas en el seno coronario.

Un criterio de referencia ideal es aquel que da siempre positivo en los pacientes que padecen la enfermedad y negativo en los que no la presentan, pero habitualmente hay que conformarse con el «mejor» disponible. Si se utiliza un mal criterio y se evalúa una nueva prueba diagnóstica que es mejor que éste, podría llegar a la conclusión de que la nueva prueba no es útil (*sesgo del criterio de referencia imperfecto*). Si la nueva prueba es más sensible que la usada como estándar, una parte de los pacientes en los que el resultado de la nueva prueba sea positivo serán clasificados como falsos positivos. Análogamente, si la nueva prueba es más específica que la estándar, una parte de los pacientes que realmente no presentan la enfermedad serán clasificados como falsos negativos.

**Ejemplo 9.8.** En un estudio que evaluaba la ecografía para el diagnóstico de litiasis biliar, se usó como estándar la colecistografía oral. En cinco pacientes, la ecografía dio un resultado positivo, mientras que la colecistografía no detectó la presencia de litiasis. Dos de estos pacientes fueron intervenidos quirúrgicamente y se halló la litiasis, por lo que, al menos en estos dos casos, la ecografía fue más exacta que la colecistografía oral.

Si no se dispone de un criterio de referencia adecuado, puede utilizarse un conjunto de pruebas para determinar la presencia o ausencia de la enfermedad. En esta situación, debe evitarse que la nueva prueba que se evalúa sea una de las que forman parte de dicho criterio compuesto, ya que su validez podría aparecer artificialmente aumentada (*sesgo de incorporación*).

**Ejemplo 9.9.** Supongamos que, para evaluar la utilidad de la determinación del colesterol total como predictor de riesgo coronario, se utiliza como criterio de referencia (medida de riesgo coronario) el índice aterogénico (cociente entre el colesterol total y el colesterol ligado a las lipoproteínas de alta densidad). Se observa que el colesterol total es muy buen predictor, ya que el propio colesterol total sirve para calcular el índice aterogénico, que es el criterio de referencia.

En el caso de no disponer de un criterio de referencia adecuado, ya sea único o un índice compuesto, debe replantearse el estudio y evaluar la concordancia entre dos métodos diagnósticos, y no la validez de uno respecto a otro.

## Aplicación de las pruebas a los sujetos

Ambas pruebas deben aplicarse a todos los sujetos. En algunos estudios puede observarse que no se ha hecho así. Cuando se utiliza la estrategia de muestreo en función de los resultados de la prueba, puede ocurrir que no se aplique el procedimiento de referencia a todos los sujetos porque la decisión de hacerlo está condicionada por los resultados de la prueba que se estudia, bien porque sea tan invasivo que se aplique sólo en los casos más graves o bien porque no sea bien tolerado por los pacientes. En esta situación se produce el llamado *sesgo de verificación o de confirmación (work-up bias)* debido a que no se obtiene la confirmación diagnóstica en todos los casos, lo que conduce a una inadecuada valoración de las características operacionales de la nueva prueba.

**Ejemplo 9.10.** El sesgo de verificación diagnóstica aparecería, por ejemplo, si los sujetos en los que se sospechara enfermedad coronaria y tuvieran una prueba de esfuerzo positiva fueran sometidos a una coronariografía (criterio de referencia) con mayor frecuencia que los que tuvieron una prueba de esfuerzo negativa. Aunque ello no sería sorprendente –ya que los clínicos son más reticentes a someter a los riesgos de una coronariografía a sujetos con una baja probabilidad de enfermedad coronaria–, el resultado de este estudio conduciría a una sobrevaloración de la utilidad de la prueba de ejercicio para diagnosticar la enfermedad. En una situación extrema, si la coronariografía sólo se realizará a los sujetos con prueba de esfuerzo positiva, sólo podrían existir individuos con la enfermedad diagnosticada por coronariografía entre los que tuvieron un resultado de la prueba positivo, por lo que la sensibilidad sería del 100%. De forma similar, sólo existirían no enfermos (es decir, con coronariografías negativas) entre los sujetos con prueba de esfuerzo positiva, ya que solamente en ellos se habría aplicado la prueba de referencia, por lo que la especificidad sería del 0%. Este ejemplo extremo muestra cómo la existencia de un sesgo de verificación puede invalidar la estimación de las características de la prueba.

© Elsevier. Fotocopiar sin autorización es un delito.

Este problema puede evitarse, al menos de forma parcial, con un seguimiento de todos los sujetos a

los que no se ha podido aplicar el procedimiento de referencia, o una muestra de ellos, con la finalidad de observar si aparece la enfermedad en estudio, y detectar así errores de clasificación.

El criterio de referencia debe aplicarse de forma ciega para que el evaluador no tenga conocimiento del resultado de la prueba que se está evaluando y evitar así sesgos de información. Del mismo modo, la aplicación de la prueba y la interpretación de los resultados deben ser ciegos e independientes del criterio de referencia. Los resultados obtenidos no deben conducir a verificaciones en caso de discordancia con el criterio de referencia, ya que se produciría un sesgo aumentando la sensibilidad y la especificidad de la prueba. Las técnicas de enmascaramiento impiden que se produzca un *sesgo de sospecha diagnóstica*.

Otra posibilidad de error aparece cuando los resultados de la prueba son dudosos o están en el límite de la normalidad. Al calcular la sensibilidad y la especificidad (anexo 3), si no se incluyen estos resultados indeterminados en el denominador, pueden obtenerse estimaciones sesgadas de estos índices.

## ESTUDIOS DE CONCORDANCIA

Estos estudios se realizan con la finalidad de evaluar si un aparato de medida o un cuestionario, por ejemplo, son fiables, o si dos observadores concuerdan en la medición de una variable.

La medición de un fenómeno está sujeta a diferentes fuentes de variación, a menudo difíciles de diferenciar en la práctica: individual, del instrumento y del observador. Cuando se evalúa la fiabilidad de una medida, deben estudiarse diferentes aspectos:

- *Repeticibilidad de la medida.* Su evaluación requiere aplicar el método de medida de la misma manera y a los mismos sujetos en dos o más momentos del tiempo (fiabilidad test-retest).
- *Concordancia intraobservador.* Su evaluación requiere que un mismo observador valore en dos o más ocasiones a los mismos sujetos.
- *Concordancia interobservador.* Su evaluación requiere que dos o más observadores valoren una misma prueba en una muestra de sujetos.

Una medida ha de ser reproducible, es decir, debe producir el mismo resultado antes de poderla comparar con otro método. Igualmente, hay que evaluar si un observador es consistente consigo mismo antes de evaluar la concordancia entre observadores.

En un estudio que evalúa la fiabilidad no siempre se estudian todas las posibles fuentes de variación.

Por ejemplo, al evaluar la fiabilidad de un método de medición automática de la presión arterial, en el que el propio aparato proporciona la cifra de presión arterial, no tiene sentido evaluar la concordancia intraobservador e interobservador. De la misma forma, en un estudio que evalúe la fiabilidad de una prueba invasiva con riesgos potenciales para el paciente, no podría evaluarse la repetibilidad, ya que no sería ético realizar la prueba en más de una ocasión.

Los sujetos incluidos en el estudio deben ser una muestra representativa de la población a la que se piensa aplicar la medida y cubrir un amplio rango de valores de la variable que se desea medir, para poder evaluar la fiabilidad durante todo su recorrido.

En el anexo 4 se presentan los índices que evalúan la concordancia entre medidas.

## Repetibilidad

La evaluación de la repetibilidad tiene por objetivo determinar si una prueba da los mismos o similares resultados cuando se aplica a una misma persona en más de una ocasión. Requiere que las condiciones de aplicación sean iguales en todas las ocasiones.

El hecho de aplicar la misma prueba en más de una ocasión en la misma muestra de individuos conlleva algunos problemas. En primer lugar, algunas de las características pueden variar con el tiempo. Es el caso de la medida del estado físico o de la salud de una persona; cuanto más tiempo transcurra entre ambas pruebas, mayor es la probabilidad de que hayan ocurrido cambios reales en su estado de salud, o de que su opinión o actitud hacia un problema de salud haya cambiado. En esta situación, la fiabilidad de la prueba será infravalorada. Esta dificultad se puede subsanar, al menos parcialmente, efectuando la segunda medición poco tiempo después de la primera. Sin embargo, si el intervalo de tiempo entre la aplicación de las dos pruebas es corto, es posible que el resultado obtenido en la segunda ocasión esté influido por el aprendizaje adquirido la primera vez, en cuyo caso ambas mediciones no serán independientes y el coeficiente de fiabilidad estará artificialmente elevado.

Cabe la posibilidad de que los participantes no accedan a que se les efectúe una prueba en más de una ocasión en poco tiempo, sobre todo cuando sea dolorosa o molesta.

## Concordancia intraobservador e interobservador

La *concordancia intraobservador* se refiere al grado de consistencia de un observador consigo mismo al leer o interpretar los resultados de dos medidas

independientes de un mismo fenómeno. Esto puede de ser fácil de conseguir cuando se interpreta una radiografía o un electrocardiograma, por ejemplo, pero es más difícil cuando se trata de un hallazgo físico que requiere la presencia directa del paciente, ya que el recuerdo de la primera exploración puede condicionar el resultado de la segunda.

Por *concordancia interobservador* se entiende la consistencia entre dos observadores independientes sobre una medida practicada en el mismo individuo. Para asegurar la independencia de las medidas entre ambos observadores, ninguno de ellos debe conocer el resultado proporcionado por el otro.

Con frecuencia, la concordancia intraobservador e interobservador se evalúa en un mismo estudio. En este caso, hay que asegurar la independencia de todas las medidas, para lo que puede ser muy útil usar técnicas como la aleatoriedad en la secuencia de la aplicación de los procedimientos de medida y las técnicas de enmascaramiento.

## Concordancia entre métodos

La concordancia entre métodos se refiere a la capacidad de dos procedimientos diferentes que miden la misma variable de dar resultados similares cuando se aplican a los mismos sujetos.

Hay que evaluar o conocer previamente la repetibilidad de cada uno de los métodos, ya que, si en uno de ellos es baja, el grado de concordancia entre ambos será probablemente bajo, aunque uno de ellos sea mucho mejor que el otro. Si ambos tienen escasa repetibilidad, el problema es aún más grave.

## ESTUDIOS ECOLÓGICOS

En los estudios ecológicos, las unidades de análisis son agregaciones de individuos, a menudo basadas en criterios geográficos o temporales. Son estudios rápidos, económicos y fáciles de realizar, especialmente si la información que se pretende recoger está disponible en anuarios estadísticos o grabada en soporte magnético.

Se pueden distinguir diferentes tipos de estudios ecológicos. En primer lugar, los estudios que comparan una medida de la frecuencia de un determinado problema (incidencia, prevalencia, mortalidad, utilización de servicios, etc.) en varias áreas, buscando la detección de algún patrón de tipo geográfico. Por ejemplo, la comparación de la mortalidad por enfermedades cardiovasculares en las comarcas de una comunidad autónoma o el atlas de mortalidad por cáncer en España. La observación de algunas diferencias según un patrón geográfico

puede dar lugar a la generación de hipótesis sobre factores sociales, culturales, económicos, de hábitos de vida, medioambientales e incluso genéticos.

En segundo lugar, los *estudios de series temporales*, que describen las variaciones de la frecuencia de un problema de salud a lo largo del tiempo, buscando patrones estacionales o tendencias (p. ej., el análisis de las tendencias de mortalidad cardiovascular en España en los últimos 40 años). Pueden ser útiles para generar hipótesis predictivas sobre la evolución de la frecuencia de un problema de salud y, por tanto, de necesidades de atención sanitaria en el futuro, o para evaluar las intervenciones sanitarias y los programas de salud.

**Ejemplo 9.11.** El brusco descenso en las tasas de morbilidad declarada por poliomielitis en España en 1964 fue la mejor evaluación de la efectividad de la campaña masiva de vacunación llevada a cabo en 1963. De todas formas, los resultados deben interpretarse con cautela, ya que si el cambio no es brusco y no se produce en poco tiempo, es difícil atribuir el descenso a una determinada intervención específica (Salleras, 1989).

Los cambios en las tendencias pueden deberse a factores distintos de la intervención, como los siguientes:

- *Cambios en las técnicas diagnósticas*, lo que conduce a una mayor sospecha y un mayor registro.

**Ejemplo 9.12.** En un estudio, en el que se analizaban las características epidemiológicas de la hidatidosis humana en Asturias durante el período 1975-1984 (Pozo et al., 1987), se recogió información de todos los hospitales asturianos. Se observó que la tasa de incidencia aumentaba considerablemente en los últimos 5 años del período. Los autores señalaban que este aumento podía deberse al acceso y empleo de las técnicas diagnósticas más recientes, que probablemente permitían descubrir más fácilmente la enfermedad, sin descartar un aumento real de su incidencia.

- *Cambios en la estructura demográfica*, que alterarían las tasas brutas o crudas, pero no las tasas específicas por edad y/o sexo.
- Si se comparan prevalencias, hay que tener en cuenta posibles *cambios en el tratamiento que disminuyan la mortalidad*, con lo que la proporción de personas con la enfermedad iría en aumento, aunque la incidencia fuera la misma.
- *Cambios en la situación socioeconómica, las características demográficas o en los estilos de vida* de las poblaciones, que pueden contribuir a la disminución de las tasas.

**Ejemplo 9.13.** Aunque la profilaxis de la fiebre reumática, basada en el correcto diagnóstico y tratamiento de la faringoamigdalitis estreptocócica, ha sido crucial en el descenso de la enfermedad, conviene no olvidar que el declinar de la fiebre reumática en los países desarrollados se inició mucho antes de que se dispusiera de antibióticos eficaces. Parece indudable que la mejoría en la calidad de vida, la mejor nutrición y el menor hacinamiento, junto con una atención sanitaria de más calidad y más accesible, han desempeñado un papel decisivo en el descenso de la fiebre reumática, sin que ninguno de ellos de manera aislada sea capaz de explicar la situación actual (Batlle, 1987).

En los *estudios de correlación ecológica* se obtiene, de cada una de las unidades de análisis, una medida sintética de la frecuencia de la enfermedad (p. ej., la incidencia, la prevalencia o la mortalidad) y una medida sintética de la frecuencia de la exposición a uno o varios factores. Las unidades de estudio pueden ser, por ejemplo, diferentes países, regiones dentro de un mismo país o comarcas de una región. El análisis se centra en determinar si las unidades ecológicas con alta frecuencia de la enfermedad también tienen tendencia a un mayor grado de exposición. Estos estudios son útiles cuando no se dispone de información individual. Esta correlación puede realizarse sobre datos referidos a un mismo momento del tiempo (p. ej., relacionar la mortalidad por infarto agudo de miocardio y el consumo de cigarrillos *per cápita* en los países europeos en un año determinado), o bien sobre tendencias temporales y comparar la tendencia en el tiempo de una enfermedad y una exposición, analizando la evolución conjunta de ambas variables (p. ej., la mortalidad por infarto agudo de miocardio y el consumo de cigarrillos en España en los últimos 30 años).

La característica fundamental de los estudios ecológicos es que no se dispone de información sobre la exposición y la enfermedad a nivel individual. Por ello, el solo hecho de que exista una asociación ecológica no garantiza que también exista de forma individual. La elaboración de conclusiones individuales inadecuadas, basada en datos ecológicos, se denomina *falacia ecológica*.

Las ventajas de rapidez, facilidad y economía de esfuerzos pueden convertirse en una amenaza. El uso de datos ya existentes hace posible que se comparén múltiples variables con la frecuencia de una enfermedad, hasta encontrar alguna que sea significativa, pero sin poder explicar un mecanismo plausible del resultado.

A pesar de sus limitaciones, los estudios ecológicos pueden ser útiles para generar hipótesis que deben ser evaluadas posteriormente con estudios

analíticos. Además, en algunas situaciones, el enfoque ecológico puede ser el más razonable. Si la hipótesis de trabajo que existe es poco firme o está poco madura, puede no ser conveniente realizar un estudio analítico individual y sea más razonable

llevar a cabo uno ecológico. También, si el interés se centra en fenómenos sociales o culturales, los estudios ecológicos pueden ser los más adecuados, ya que se pretende obtener inferencias sobre comunidades completas más que sobre individuos.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- Bates SE. Clinical applications of serum tumor markers. *Ann Intern Med.* 1991;115:623-38.
- Batle Gualda E. ¿Dónde está la fiebre reumática? *Med Clin (Barc).* 1987;88:61-4.
- Davis M, Ettinger W, Neuhaus J, Hauck W. Sex differences in osteoarthritis of the knee: the role of obesity. *Am J Epidemiol.* 1988;127:1019-29.
- Melton J, Ochi J, Palumbo P, Chu Pin Chu. Sources of disparity in the spectrum of diabetes mellitus at Incidence and Prevalence. *Diabetes Care.* 1983;6:427-31.
- Pozo F, Fernández MJ, Suárez TV, Tojo S, Lamamie E, Rodrigo LR. Estudio epidemiológico de la hidatidosis en Asturias (1975-1984). *Med Clin (Barc).* 1987;89:773-7.
- Salleras Sanmartí L. Estudios descriptivos. *Aten Primaria.* 1989;6:504-10.
- Thomson DMP, Krupey J, Freedman SO, Gold P. The radioimmunoassay of circulating carcino-embryonic antigen of the human digestive system. *Proc Natl Acad Sci USA.* 1969;64:161-7.

## BIBLIOGRAFÍA

- Bachmann LM, Puhan MA, ter Riet G, Bossuyt PM. Sample sizes of studies on diagnostic accuracy: literature survey. *BMJ.* 2006;332(7550):1127-9.
- Cabello López JB, Pozo Rodríguez F. Estudios de evaluación de las pruebas diagnósticas en cardiología. *Rev Esp Cardiol.* 1997;50:507-19.
- Greenland S. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *Int J Epidemiol.* 2001;30:1343-50.
- Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet.* 2005;365(9469):1500-5.
- Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. *Lancet.* 2002;359:145-9.
- Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. Designing studies to ensure that estimates of test accuracy are transferable. *BMJ.* 2002;324:669-71.
- Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA.* 1999;282:1061-6.
- Rigau D, Alonso P, Solà I. Estudios de precisión diagnóstica y de pronóstico: las directrices STARD y REMARK. En: Jiménez Villa J, Argimon Pallás JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo científico. Barcelona: Elsevier; 2010. p. 215-29.
- Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ.* 2002;324: 539-41.
- Wakefield J. Ecologic studies revisited. *Annu Rev Public Health.* 2008;29:75-90.
- Willis BH. Spectrum bias—why clinicians need to be cautious when applying diagnostic test studies. *Fam Pract.* 2008;25(5):390-6.

# Capítulo

# 10

## Revisiones sistemáticas

Una revisión sistemática (RS) de la evidencia científica consiste en la síntesis de la mejor evidencia disponible para responder a una pregunta concreta mediante la aplicación de una metodología explícita y rigurosa. Para ello utiliza el mismo procedimiento que cualquier estudio de investigación: se formula un objetivo, se busca, se identifican y se valoran las pruebas disponibles siguiendo un protocolo con criterios explícitos, y se obtienen conclusiones a partir del análisis de los datos recogidos y la interpretación de los resultados. Aunque con frecuencia los términos revisión sistemática y metaanálisis se emplean como sinónimos, hoy en día se reserva el término metaanálisis para el procedimiento estadístico que se utiliza para integrar los resultados de diferentes estudios y obtener una estimación conjunta, mientras que el concepto de RS se aplica a todo el proceso de la investigación descrito anteriormente.

Las RS se han convertido en una herramienta de gran utilidad en el proceso de toma de decisiones en la práctica clínica, por lo que es imprescindible que se realicen con el máximo rigor. Dada su gran trascendencia, y con la finalidad de mejorar la calidad de su publicación en las revistas médicas, se han elaborado unas directrices –conocidas como PRISMA– sobre la información que debe incluirse en un artículo que presente una RS (Liberati et al, 2009).

### FASES DE UNA REVISIÓN SISTEMÁTICA

Una RS es un estudio de investigación que, a diferencia de los presentados en capítulos anteriores, no se realiza sobre datos primarios, es decir, no recoge la información directamente de los sujetos,

sino que utiliza los datos recogidos previamente en otros estudios. Una RS puede considerarse un estudio observacional en que la «población de estudio» está formada por los mejores artículos originales realizados sobre el tema objeto de la revisión. Como cualquier otro estudio, requiere elaborar un protocolo que especifique las definiciones y los procedimientos que se efectuarán en sus distintas etapas (cuadro 10.1).

### DEFINICIÓN DEL OBJETIVO

La finalidad de una RS es intentar responder una pregunta clínica concreta valorando todas las pruebas disponibles procedentes de las investigaciones realizadas. El primer paso es, pues, definir con claridad y precisión dicha pregunta, explicitando el factor de estudio (intervención, exposición o procedimiento diagnóstico que se evalúa) y el de comparación, la variable de respuesta (especificando la medida del efecto que se utilizará, como la reducción de la incidencia o la mortalidad, la mejoría de síntomas, etc.) y la población de estudio (especificando el tipo de paciente y la condición clínica). También conviene que se identifique el tipo de estudio que se incluirá. En general, cuanto más precisa sea la definición de la pregunta que se desea contestar, mejor será la RS.

### IDENTIFICACIÓN DE LOS ESTUDIOS DISPONIBLES

Para que las conclusiones de una RS sean válidas, hay que identificar y localizar todos los estudios existentes, idealmente tanto los publicados como

### Cuadro 10.1 Guía para la realización de una revisión sistemática

- Definición del objetivo (pregunta de investigación)
- Búsqueda de evidencias:
  - Bases de datos consultadas
  - Palabras clave utilizadas
  - Período de cobertura
  - Otros requisitos: revistas, idioma, etc.
- Criterios de inclusión y exclusión de estudios
- Evaluación de la calidad de los estudios:
  - Resumen de las características esenciales de los estudios
  - Calificación de la calidad de los estudios
  - Evaluación de la variabilidad entre los investigadores que califican la calidad de los estudios
- Recogida de datos:
  - Registro de las características de los estudios:
    - Tipo de artículo y año de publicación
    - Diseño del estudio
    - Características de la intervención
    - Características del grupo control
    - Tamaño de la muestra
  - Registro de los resultados de los estudios
- Análisis de los resultados:
  - Pruebas de homogeneidad
  - Combinación estadística de los resultados
  - Pruebas para detectar diferencias sistemáticas entre estudios
  - Representaciones gráficas
  - Análisis de sensibilidad
  - Análisis de subgrupos
- Conclusiones y recomendaciones

los no publicados. Habitualmente, la búsqueda de evidencias se realiza en una o varias de las bases de datos más importantes (MEDLINE, EMBASE, Biblioteca Cochrane), pero es conveniente complementarla con una búsqueda manual en las referencias de los artículos localizados, libros y otras revisiones, y también a través de contactos con expertos en la materia para aumentar la exhaustividad de la detección de los estudios e intentar identificar trabajos no publicados.

El protocolo de la RS debe especificar las bases de datos que se consultarán, las palabras clave y la estrategia de búsqueda que se utilizarán, el período de tiempo de cobertura y cualquier otro requisito que se piense utilizar en ella.

### CRITERIOS DE SELECCIÓN

Los estudios identificados forman la población candidata. La selección de los que se incluirán en la RS debe hacerse a partir de criterios explícitos y establecidos con claridad, de forma que el proceso sea reproducible por otros investigadores.

Los criterios de selección deben referirse a características fundamentales de los estudios. El objetivo fundamental es obtener un conjunto de trabajos que aborden una misma pregunta científica y con la calidad suficiente para que sus resultados puedan considerarse válidos. Estos criterios deben ser fácilmente objetivables, ya que habitualmente su cumplimiento se evaluará mediante la lectura detallada del artículo, por lo que deben basarse en información fácilmente accesible en él y que se refleje de forma explícita.

En primer lugar, se define el tipo de estudio y sus características principales, y a continuación se establecen otros criterios relacionados habitualmente con su calidad metodológica. Los más frecuentemente utilizados se refieren a:

- *Las intervenciones o exposiciones que se evalúan (factor de estudio).* Obviamente, si se desea integrar los resultados de diferentes trabajos, éstos deben evaluar un mismo factor de estudio.
- *La variable de respuesta.* Si los estudios utilizan diferentes criterios de evaluación o diferentes variables de respuesta, resultará muy difícil poder integrar sus resultados.
- *Criterios de selección de la población de los estudios originales.* El efecto puede ser diferente en función de diversas características de los sujetos estudiados. Por ello, pueden utilizarse criterios muy restrictivos para incluir sólo estudios que evalúen poblaciones homogéneas, o bien más laxos y utilizar las diferentes características para realizar estratificaciones y análisis de subgrupos.
- *Tamaño muestral mínimo* para asegurar una precisión suficiente en la estimación de los resultados. A veces se decide excluir los estudios muy pequeños, ya que pueden estar realizados de forma poco rigurosa y sin controles de calidad adecuados.
- *Duración del seguimiento.* Es necesario asegurar que los estudios han realizado un seguimiento suficiente para poder observar la respuesta de interés.
- *Exhaustividad de la información necesaria presentada en el artículo original.*

## VALORACIÓN CRÍTICA DE LOS ESTUDIOS SELECCIONADOS

Idealmente, sólo deberían incluirse en una RS aquellos estudios con una calidad mínima suficiente, establecida de antemano, para asegurar que sus resultados puedan considerarse válidos. La evaluación de la calidad de un estudio es un proceso subjetivo, y más teniendo en cuenta que en ocasiones la información necesaria no está disponible o es inadecuada.

Existen múltiples escalas para realizar esta valoración. En la tabla 10.1 se presenta una de las más sencillas y más utilizadas.

La mayoría de las escalas conceden una puntuación de calidad a cada uno de los estudios, que puede utilizarse de diferentes maneras: para definir un punto de corte a partir del cual identificar los artículos que van a ser incluidos o excluidos de la RS, para ponderar el peso relativo de cada uno de los estudios en el análisis, para evaluar si existe relación entre la magnitud del efecto observado y las puntuaciones de calidad, etc. Esta última finalidad podrá ser realizada mediante un método gráfico, como un metaanálisis acumulativo en que los estudios sean incluidos sucesivamente en el análisis según su puntuación de calidad, de la mayor a la menor, o bien mediante un procedimiento estadístico (metarregresión) que permite analizar la posible relación entre el efecto observado y determinadas covariables relacionadas con la calidad, como el tipo de diseño (ensayo clínico, estudio de cohortes, estudio de casos y controles), la procedencia de los datos (entrevistas, cuestionarios por correo, etc.) o la procedencia de los sujetos (registros poblacionales, hospitales, etc.).

Es conveniente registrar el cumplimiento de los criterios de selección en todos los estudios

candidatos, ya que es recomendable prever análisis de sensibilidad, repitiendo el metaanálisis en función de diferentes criterios de inclusión y de exclusión para examinar si el resultado final y las conclusiones de la RS cambian en función de dichos criterios.

## EXTRACCIÓN DE LOS DATOS

El protocolo de la RS ha de contemplar la información que se extraerá de los diferentes estudios: datos generales del estudio, nombre del autor, año de realización, características principales del diseño, tipo de pacientes (características demográficas, condición clínica, etc.), características metodológicas (tamaño de la muestra, enmascaramiento, duración del estudio, etc.), características de las intervenciones de estudio y de comparación (pauta de intervención, duración, etc.), comparabilidad inicial de los grupos, variables de respuesta, etc.

La recogida de información sobre el resultado del estudio es fundamental. Debe utilizarse un único –o, como mínimo, que sea comparable– sistema de clasificación del evento de interés (muerte, aparición de enfermedad, mejoría de los síntomas, etc.). Se debería escoger también el mismo tipo de estimador para todos los estudios, o calcularlo de nuevo si no aparece en el artículo, por lo que éste deberá contener la información necesaria. De no ser así, habrá que contactar con los investigadores para tratar de obtenerla.

Es conveniente que la valoración crítica y la extracción de datos sean realizadas como mínimo por dos revisores independientes y, a ser posible, que desconozcan los resultados, los autores y la revista donde se publicaron los artículos primarios. Incluso aunque los investigadores sigan estrictamente un protocolo de recogida de datos con criterios explícitos y bien definidos suele existir cierto

**Tabla 10.1 Escala de Jadad para valorar la calidad de un ensayo clínico aleatorio**

Pregunta	Puntos	Puntos adicionales
¿Se indica si el estudio fue aleatorio?	1: Sí 0: No	+1: Se describe el método de aleatorización y es adecuado -1: Se describe el método de aleatorización y es inadecuado
¿Se indica si el estudio fue doble ciego?	1: Sí 0: No	+1: Se describe el método de enmascaramiento y es adecuado -1: Se describe el método de enmascaramiento y es inadecuado
¿Hay una descripción de las pérdidas de seguimiento?	1: Sí 0: No	

Puntuación total: de 0 a 5 puntos (a más puntuación, mayor calidad).

componente de subjetividad. Es importante evaluar la concordancia entre los observadores para asegurar la calidad de los datos que se utilizarán.

## ESTRATEGIA DE ANÁLISIS

El último paso de una RS suele ser la combinación estadística de los distintos estudios (metaanálisis), si es posible, para obtener un resultado global que estime el efecto de la intervención, exposición o prueba diagnóstica en estudio. Las ventajas potenciales del metaanálisis incluyen el aumento de la potencia estadística y de la precisión de los resultados, así como la capacidad para responder preguntas no abordadas por los estudios individuales. Sin embargo, los procedimientos estadísticos no pueden suplir la falta de sentido común o de criterio clínico en el diseño del protocolo de la revisión, o de plausibilidad biológica en la interpretación de los resultados. Así pues, el lector de una RS debe evaluar estos aspectos antes de examinar los métodos estadísticos que han usado los revisores y los resultados que han obtenido. La combinación de datos de poca calidad, muy sesgados o procedentes de estudios muy dispares entre sí producirá un resultado poco creíble.

En el cuadro 10.2 se presenta el esquema de la estrategia de análisis de una RS. En primer lugar, debe evaluarse si la validez de los estudios es suficiente y si tiene sentido integrar los diferentes estudios, para decidir si es adecuado combinar estadísticamente

sus resultados. Después, debe decidirse qué comparaciones se van a realizar, definidas con claridad a priori y de acuerdo con el objetivo de la RS.

El primer paso es analizar la heterogeneidad. Solamente tiene sentido combinar estadísticamente los resultados si los estudios son lo suficientemente homogéneos. En caso contrario, debe explorarse esta heterogeneidad y tratar de identificar los factores de los que depende y que pueden explicarla.

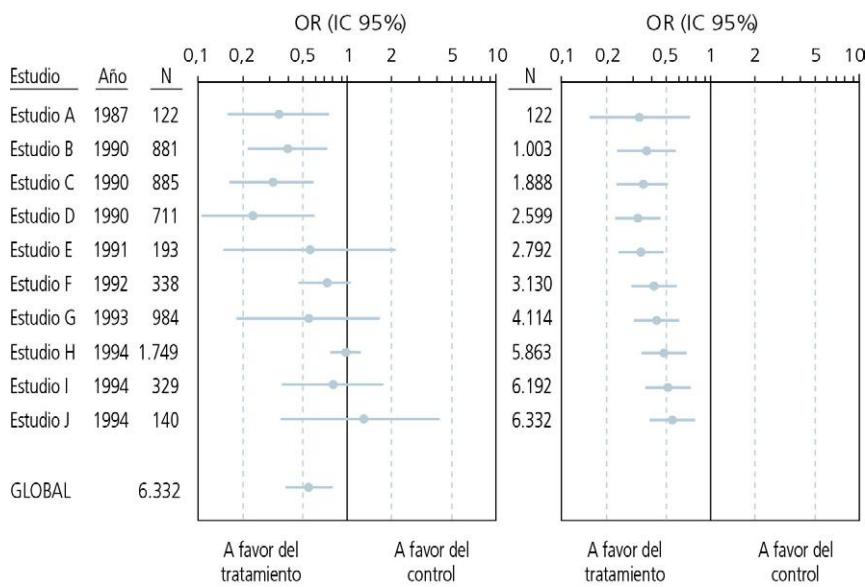
Si se puede realizar un metaanálisis y combinar los resultados, debe determinarse qué tipo de medida se utilizará para obtener una estimación global del efecto y qué procedimiento estadístico se aplicará para calcular su magnitud y su intervalo de confianza. Los resultados de este metaanálisis suelen presentarse gráficamente para facilitar la comprensión (fig. 10.1). Habitualmente, el resultado de cada uno de los estudios individuales se muestra como una línea horizontal que representa el intervalo de confianza (IC) del 95% de la estimación de la medida del efecto, con una señal que indica la estimación puntual observada en dicho estudio. Una última línea horizontal representa el resultado global obtenido con el metaanálisis, cuya estimación puntual suele representarse por un rombo. Una línea vertical representa la ausencia de efecto (valor 1 cuando se utilizan medidas relativas, como la *odds ratio* o el riesgo relativo, o valor 0 cuando se utilizan medidas absolutas, como la diferencia de incidencias). Cuando los estudios se presentan ordenados por algún criterio, como el año de su realización o la puntuación de calidad, como ocurre en el gráfico de la derecha de la figura 10.1, se habla de metaanálisis acumulativo.

También debe planificarse si se realizarán análisis de subgrupos para determinar la magnitud del efecto en subpoblaciones de diferentes características, así como los análisis de sensibilidad, dirigidos a evaluar hasta qué punto los resultados del metaanálisis dependen de determinadas decisiones tomadas en el diseño de la RS relacionadas con los criterios de inclusión, la inclusión de estudios de baja calidad o no publicados, etc. En el anexo 5 se presenta una descripción más detallada del análisis de una RS.

La finalidad del análisis de una RS es evaluar el conjunto de las pruebas disponibles y ayudar a interpretarlas correctamente para intentar responder a la pregunta planteada, teniendo en cuenta también sus implicaciones para la práctica clínica y el grado de aplicabilidad a diferentes entornos. Uno de los instrumentos más utilizados para evaluar la calidad de dichas pruebas es el desarrollado por el Grades of Recommendation, Assessment, Development and Evaluation Working Group (**GRADE** Working Group, 2004; Guyatt et al, 2008).

Cuadro 10.2 Fases de la estrategia de análisis de una revisión sistemática

- Decidir si es adecuado realizar una integración de los estudios
- Decidir qué comparaciones se van a hacer
- Analizar la heterogeneidad de los estudios (tablas de evidencias, pruebas estadísticas, métodos gráficos)
- Si es adecuado, calcular un estimador global del efecto (metaanálisis) con su intervalo de confianza
- Presentación gráfica de los resultados
- Análisis de subgrupos
- Análisis de sensibilidad:
  - Para evaluar la robustez del resultado
  - Para explorar la heterogeneidad de los estudios
- Evaluar la existencia de un posible sesgo de publicación
- Pruebas para detectar diferencias sistemáticas entre estudios



**Figura 10.1** Presentación gráfica de una misma revisión sistemática. En el gráfico de la derecha se presentan los resultados en forma acumulativa.

## SESGOS EN UNA REVISIÓN SISTEMÁTICA

En una RS pueden aparecer sesgos en diferentes fases, especialmente en las de localización y selección de los estudios; algunos de estos sesgos pueden amenazar la validez de las conclusiones.

### Sesgo de publicación

El sesgo de publicación se refiere al hecho de que no todos los estudios tienen la misma probabilidad de ser publicados. Con frecuencia, los autores deciden no enviar sus manuscritos a las revistas, o bien los editores y los revisores de una revista deciden no aceptar algunos estudios en función de determinadas características, relacionadas más con los resultados encontrados que con aspectos de calidad. Ello hace que los estudios multicéntricos, los que cuentan con financiación externa y los que encuentran resultados estadísticamente significativos, se publiquen con mayor frecuencia, introduciendo un sesgo en la evaluación de las pruebas disponibles si sólo se toman en consideración las publicadas.

La prevención de este sesgo es importante desde dos perspectivas: en primer lugar, la científica, para conseguir una completa difusión del conocimiento; si las decisiones de los profesionales se han de basar en los datos publicados, la literatura debe incluir todos los

estudios que tengan una calidad aceptable. En segundo lugar, desde la perspectiva de los autores que realizan una RS, si se publican preferentemente artículos con resultados positivos, cualquier RS tenderá a obtener también resultados positivos (fig. 10.2).

### Sesgo en la localización de estudios

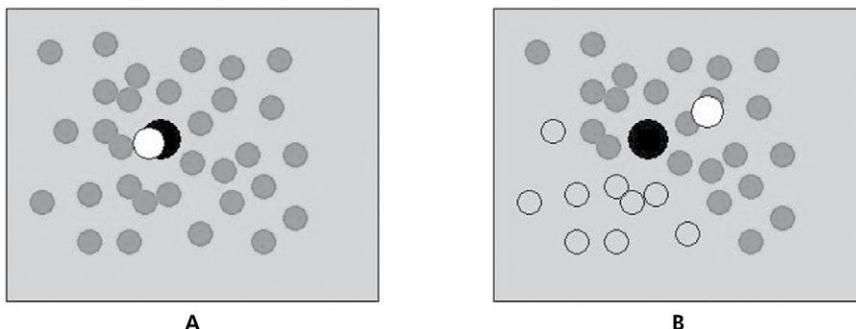
Además del sesgo de publicación, existen otros muchos factores que contribuyen a una inclusión sesgada de estudios en una RS. Los más importantes son el sesgo del idioma inglés, el sesgo de las bases de datos, el sesgo de citación y el sesgo de publicación múltiple.

### Sesgo del idioma inglés

Muchas RS sólo incluyen artículos publicados en inglés, de manera que los estudios publicados en otros idiomas quedan excluidos. Además, no es infrecuente que los investigadores de países de habla no inglesa publiquen los estudios con resultados positivos en revistas de lengua inglesa, al considerarlas más relevantes, y los negativos en una revista local, con lo que este sesgo positivo se añade al inherente al de publicación.

### Sesgo de las bases de datos

Las dos bases de datos bibliográficos más utilizadas, MEDLINE y EMBASE, no tienen la misma cobertura y, por lo tanto, si la búsqueda se limita



**Figura 10.2** Sesgo de publicación. Cada círculo representa el resultado de un estudio. El círculo negro representa el resultado global de los estudios incluidos en una revisión sistemática. El círculo blanco representa la verdadera magnitud del efecto. A: revisión sistemática que ha incluido todos los estudios existentes. Su resultado se aproxima al verdadero efecto. B: revisión sistemática en que algunos estudios no han sido incluidos. Su resultado está sesgado.

a artículos indizados solamente en una de ellas, puede introducirse un sesgo. Además, los criterios de selección a la hora de indizar las revistas hacen que muchas de ellas, publicadas en determinados países, no lo estén. Por ejemplo, de las más de 4.000 revistas indizadas en EMBASE, MEDLINE o Science Citation Index, sólo el 2% procede de países en vías de desarrollo.

Los artículos que se publican en revistas no indizadas en estas bases de datos difícilmente serán localizados por los autores de una RS, y si sus resultados difieren del obtenido en los estudios indizados, el resultado de la revisión estará sesgado. De hecho, es más probable que los artículos con resultados positivos estén publicados en revistas indizadas.

### Sesgo de citación

Para intentar localizar los estudios que deben incluirse en una RS es habitual que los autores complementen la búsqueda en las bases de datos contactando con expertos en la materia y realizando una búsqueda manual a partir de las referencias bibliográficas de los estudios publicados. Cuando se efectúa esta búsqueda manual es probable que se introduzca un sesgo de citación, ya que, por regla general, los estudios con resultados positivos se citan con más frecuencia que los estudios con resultados negativos. Por ello, es más probable incluir en una RS estudios con resultados positivos y sesgar su resultado.

La revista en la que se publica un estudio también puede influir en su citación y, por tanto, en su probabilidad de que dicho estudio sea incluido en una revisión.

**Ejemplo 10.1.** En el estudio de Helsinki, un ensayo clínico en el que se evaluó la eficacia de un fibrato para prevenir la cardiopatía isquémica en pacientes sin antecedentes de la enfermedad, se obtuvieron resultados estadísticamente significativos y fue publicado en una prestigiosa revista (Frick et al, 1987). El estudio tenía otra parte destinada a evaluar la eficacia de la intervención para la prevención secundaria de la enfermedad, cuyos resultados no fueron estadísticamente significativos. Los resultados de esta segunda parte se publicaron 6 años después en una revista de mucho menor impacto (Frick et al, 1993). El primer artículo fue citado en más de 450 ocasiones durante los 3 años siguientes a su publicación, mientras que el segundo fue citado sólo en 17 ocasiones.

### Sesgo de publicación múltiple

La publicación múltiple de resultados de un mismo estudio es un fenómeno frecuente que se da especialmente en los estudios de gran envergadura, y que puede conducir a un sesgo. Por un lado, los estudios con resultados estadísticamente significativos tienden a publicarse con más frecuencia; por lo tanto, es más fácil localizarlos e incluirlos en una revisión. Por otro lado, si no se identifican como publicación múltiple, pueden incluirse datos duplicados que conduzcan a una sobreestimación del efecto. En ocasiones es difícil deducir que los resultados publicados en diferentes artículos proceden de un mismo estudio, especialmente en estudios multicéntricos en los que, además de los resultados globales, se publican resultados de centros aislados.

## Sesgo por datos ausentes o de mala calidad

En ocasiones, los artículos no incluyen todos los datos necesarios para efectuar una RS. Muchos factores pueden estar relacionados con la voluntad de los investigadores de hacer públicos estos resultados, pero uno de ellos puede ser la dirección y la significación estadística de los mismos. Este sesgo es más probable en los estudios publicados como comunicaciones a congresos y en los no publicados.

## Sesgo en la definición de los criterios de inclusión

También es posible introducir un sesgo cuando se establecen los criterios de inclusión y exclusión, ya que son definidos por un equipo investigador familiarizado con el tema y, por consiguiente, con los principales estudios publicados. Este conocimiento puede influir a la hora de definir estos criterios, lo que puede conducir a una inclusión selectiva de determinados estudios.

## REVISIONES SISTEMÁTICAS QUE USAN DATOS DE PACIENTES INDIVIDUALES

Las RS se basan habitualmente en los datos publicados en los artículos originales. Sin embargo, en ocasiones, la presentación de los datos no favorece su revisión. Por ejemplo, en una RS sobre la eficacia del tratamiento antihipertensivo en ancianos puede ocurrir que algún estudio no incluya a los pacientes mayores de 60 años, mientras que otros establezcan un límite de edad distinto. Además, como muchos estudios no presentan resultados estratificados por edad, sus datos no podrán incluirse en el análisis. La recogida de los datos de los pacientes individuales directamente de los investigadores puede resolver alguno de estos problemas, ya que se solicitan datos sobre tipos específicos de pacientes y resultados.

Los principios que guían una revisión con datos de pacientes individuales son los mismos que los de cualquier RS. La principal diferencia es que incorpora, en la medida de lo posible, toda la evidencia relevante disponible. Se trata de conseguir que todos los pacientes aleatorizados y ninguno de los no aleatorizados de cualquier ensayo considerado válido y pertinente sean incluidos y se analicen siguiendo el principio de «intención de tratar».

Este tipo de RS son el medio más fiable para combinar datos de ensayos clínicos similares. Sus principales ventajas son las siguientes:

- Se pueden realizar análisis de supervivencia, si la variable de respuesta es el tiempo de supervivencia o el tiempo hasta que se produce un suceso. Este enfoque es más potente e informa mejor sobre el curso de la enfermedad y el efecto del tratamiento en el transcurso del tiempo.
- Se pueden efectuar análisis de subgrupos con una definición común para poner a prueba y generar hipótesis.
- Se puede controlar mejor la calidad de los datos, teniendo la posibilidad de validarlos y corregirlos si fuera necesario, mediante la comunicación directa con los investigadores.
- Es más fácil detectar desviaciones en el protocolo de los estudios individuales.
- Se puede actualizar la información sobre el seguimiento de los pacientes cuando los registros están disponibles.
- Es más fácil estandarizar las medidas de resultado.

Un problema que puede surgir en estas RS es que los datos de un estudio o más de uno no estén disponibles. Puede considerarse que, si se ha logrado reunir el 90-95% de los estudios originales, los datos no disponibles no cambiarían sustancialmente los resultados de la RS.

## REVISIONES SISTEMÁTICAS DE ESTUDIOS OBSERVACIONALES

El tipo de estudios que se incluyen en la RS puede condicionar los resultados. Dado que el ensayo clínico aleatorio es el diseño que más asegura la comparabilidad de los grupos, la mayoría de las RS incluyen solamente este tipo de estudios, lo que ha permitido desarrollar consensos sobre cómo debe aplicarse esta metodología. Sin embargo, existe más controversia sobre cómo realizar una RS cuando se incluyen también estudios observacionales, ya que en estos diseños es más probable la presencia de sesgos cuya magnitud y dirección a menudo es difícil de prever.

Existen otros puntos de controversia en la realización de RS con estudios observacionales, como si es o no correcto combinar distintos tipos de diseño (p. ej., estudios de casos y controles con estudios de cohortes), si pueden combinarse estudios de casos y controles que utilizan controles poblacionales con otros que utilizan controles hospitalarios, o si es razonable combinar

estudios de cohortes procedentes de varios países o de distintos ámbitos (laboral, poblacional, etc.).

Algunos autores piensan que no deberían efectuarse nunca RS de estudios observacionales, mientras que otros sugieren que lo que no debe efectuarse

es la combinación estadística de sus resultados. La consideración cuidadosa de las posibles fuentes de heterogeneidad entre los estudios proporcionará probablemente datos más valiosos que la simple combinación estadística de los resultados.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- 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;317:1237-45.
- Frick MH, Heinonen OP, Huttunen JK, Koskinen P, Marttari M, Manninen V. Efficacy of gemfibrozil in dyslipidaemic subjects with suspected heart disease. *Ann Med.* 1993;25:41-5.

## BIBLIOGRAFÍA

- Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analysis of randomized controlled trials. *JAMA.* 2002;287:2973-82.
- Crowther M, Lim W, Crowther MA. Systematic review and meta-analysis methodology. *Blood.* 2010;116:3140-6.
- Egger M, Ebrahim S, Smith GD. Where now for meta-analysis? *Int J Epidemiol.* 2002;31(1):1-5.
- Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ.* 1998;316:140-4.
- Egger M, Smith GD. Ias in location and selection of studies. *BMJ.* 1998;316:61-6.
- Egger M, Smith GD, Sterne JA. Uses and abuses of meta-analysis. *Clin Med.* 2001;1:478-84.
- GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ.* 2004;328:1490-4.
- Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ.* 2008;336:924-6.
- Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011]. The Cochrane Collaboration, 2011. Available from: [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
- Higgins J, Thompson S, Deeks J, Altman D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *J Health Serv Res Policy.* 2002;7:51-61.
- Juni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analysis of controlled trials: empirical study. *Int J Epidemiol.* 2002;31:115-23.
- Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA.* 1999;282:1054-60.
- Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analysis. *Ann Intern Med.* 2001;135:982-9.
- Kranke P. Evidence-based practice: how to perform and use systematic reviews for clinical decision-making. *Eur J Anaesthesiol.* 2010;27:763-72.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *BMJ.* 2009;339:b2700.
- Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Stat Med.* 2001;20:641-54.
- McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analysis? *Lancet.* 2000;356:1228-31.
- Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, et al. What contributions do languages other than English make on the results of meta-analysis? *J Clin Epidemiol.* 2000;53:964-72.
- Montori VM, Smieja M, Guyatt GH. Publication bias: a brief review for clinicians. *Mayo Clin Proc.* 2000;75:1284-8.
- Mullen PD, Ramirez G. The promise and pitfalls of systematic reviews. *Annu Rev Public Health.* 2006;27:81-102.

- Palma Pérez S, Delgado Rodríguez M. Consideraciones prácticas acerca de la detección del sesgo de publicación. *Gac Sanit.* 2006;20Supl3:10-6.
- Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 13. Applicability, transferability and adaptation. *Health Research Policy and Systems.* 2006;4:25.
- Silagy CA, Middleton P, Hopewell S. Publishing protocols of systematic reviews: comparing what was done to what was planned. *JAMA.* 2002;287(21):2831-4.
- Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Med Res Methodol.* 2011;11:15.
- Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ.* 2001;323:101-5.
- Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol.* 2000;53:1119-29.
- Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in «meta-epidemiological» research. *Stat Med.* 2002;21:1513-24.
- Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analysis. *BMJ.* 2000;320:1574-7.
- Tugwell P, Knottnerus A, Idzerda L. Updating systematic reviews—when and how? *J Clin Epidemiol.* 2011;64:933-5.
- Urrutia G, Bonfill X. Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Med Clin (Barc).* 2010;135:507-11.
- Vale CL, Tierney JE, Stewart LA. Effects of adjusting for censoring on meta-analysis of time-to-event outcomes. *Int J Epidemiol.* 2002;31:107-11.
- Verhagen AP, De Vet HC, De Bie RA, Boers M, Van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews. *J Clin Epidemiol.* 2001;54:651-4.
- Weed DL. Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related. *Int J Epidemiol.* 2000;29:387-90.
- Wood L, Egger M, Gluud LL, Schulz K, Jüni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ.* 2008;336:601-5.

Página deliberadamente en blanco

# Parte

**3**

## Elaboración del protocolo de estudio

<b>11</b> Protocolo de investigación <b>109</b>	<b>19</b> Selección y definición de las variables <b>181</b>
<b>12</b> Objetivo del estudio <b>113</b>	<b>20</b> Diseño de cuestionarios <b>188</b>
<b>13</b> Variable de respuesta <b>121</b>	<b>21</b> Validación de cuestionarios <b>200</b>
<b>14</b> Población de estudio <b>135</b>	<b>22</b> Estrategia de análisis <b>211</b>
<b>15</b> Tamaño de la muestra <b>142</b>	<b>23</b> Sujetos a incluir en el análisis <b>221</b>
<b>16</b> Muestreo <b>155</b>	<b>24</b> Gestión del estudio <b>232</b>
<b>17</b> Formación de los grupos de estudio <b>163</b>	<b>25</b> Solicitud de ayuda para la financiación <b>242</b>
<b>18</b> Medición de variables <b>173</b>	

Página deliberadamente en blanco

## Protocolo de investigación

Un estudio se inicia cuando un investigador se plantea una pregunta y elabora una hipótesis de trabajo. A continuación, diseña un estudio para obtener una respuesta válida y fiable.

Todos los aspectos que tengan relación con el estudio, desde la definición del objetivo hasta la estrategia del análisis, deben ser meditados y debatidos por todo el equipo investigador, y escribirse en el protocolo del estudio. Este documento incluye dos grandes bloques de información: por un lado, la estructura científica del proyecto, donde se describe la razón de ser del estudio, sus objetivos, su diseño, su estrategia de análisis, etc.; y, por otro, el manual de instrucciones para el personal investigador, donde se especifican los métodos y procedimientos que se van a utilizar en cada uno de los sujetos incluidos. Dentro de un protocolo, ambos aspectos pueden estar mezclados o aparecer en documentos separados.

El protocolo es un instrumento básico, ya que permite pasar de la concepción de un problema de investigación a su puesta en marcha. Ayuda al investigador a definir claramente el problema de investigación, a diseñar un estudio que sea factible realizar y a organizar las actividades de forma eficiente. Facilita la comunicación dentro del equipo y la estandarización de todos los procedimientos. En definitiva, permite llevar a cabo el estudio como fue diseñado y asegurar que el paso del tiempo o los cambios en el personal no lo modificarán ni lo limitarán. Además, es imprescindible para efectuar la revisión técnica competente y la crítica constructiva previa al inicio del estudio.

### ESQUEMA DEL PROTOCOLO DE ESTUDIO

El esquema general de un protocolo está íntimamente ligado a las fases del proyecto de investigación (cuadro 11.1). Cada uno de sus apartados se expone ampliamente en los siguientes capítulos.

En primer lugar, debe identificarse el problema que se desea resolver, encuadrándolo en su marco teórico, acompañándolo de una breve pero cuidadosa revisión bibliográfica muy relacionada con él y que evite revisiones excesivamente genéricas. La revisión de la literatura permite matizar la idea original, delimitar con mayor precisión el objetivo del estudio, valorar los conocimientos actuales sobre el tema, evaluar la pertinencia y la viabilidad del proyecto, proporcionar un marco conceptual para la investigación y obtener información sobre aspectos concretos del método, como criterios de selección, instrumentos de medida y análisis estadístico, entre otros.

Una idea puede ser muy interesante, pero no abordable en un medio concreto. Hay que valorar si las variables de interés pueden medirse de forma precisa y exacta, el tiempo necesario para la investigación, la disponibilidad de sujetos para alcanzar el tamaño necesario, la disponibilidad de tiempo del equipo investigador, la posibilidad de obtener la colaboración de otros profesionales o centros, si se dispone de las instalaciones, equipos y recursos económicos necesarios, y si el equipo investigador tiene la experiencia suficiente para garantizar que el estudio se realizará con el rigor necesario.

### Cuadro 11.1 Esquema general de un protocolo de estudio

1. *Introducción:*
  - Finalidad del estudio: problema que intenta resolver y preguntas que se formulan
  - Antecedentes del tema: breve revisión bibliográfica
  - Razones por las que la investigación propuesta es importante
2. *Objetivos específicos del estudio*
3. *Aplicabilidad y utilidad de los resultados*
4. *Diseño y métodos:*
  - Tipo de diseño
  - Población de estudio:
    - Descripción de la muestra
    - Criterios de inclusión y exclusión
    - Método de muestreo
    - Cálculo del tamaño de la muestra
    - Procedencia de los sujetos
  - Método de recogida de datos: entrevista, examen físico, procedimientos de laboratorio y otras pruebas complementarias
  - Variables:
    - Definición de las variables
    - Medición de las variables
    - Registro de los valores de las variables (formularios)
  - Descripción y definición de la intervención (si la hay)
  - Descripción del seguimiento de los pacientes (si lo hay)
  - Entrada y gestión informática de los datos
  - Estrategia de análisis
5. *Calendario previsto para el estudio*
6. *Limitaciones y posibles sesgos del estudio*
7. *Problemas éticos:*
  - Método que se utilizará para obtener el consentimiento informado
  - Riesgos físicos, sociales o legales a los que pueden verse sometidos los pacientes
  - Métodos utilizados para minimizar los riesgos principales
  - Beneficios potenciales que pueden obtener los participantes en el estudio
8. *Plan de ejecución. Estudio piloto*
9. *Organización del estudio:*
  - Personal que interviene y su responsabilidad
  - Instalaciones, instrumentación y técnicas necesarias
10. *Presupuesto*
11. *Bibliografía*

Debe formularse el objetivo específico del estudio (o los objetivos, si hay más de uno). El objetivo es el eje en torno al cual gira el diseño del estudio y es la base para decidir los criterios de selección de los sujetos y las variables que han de ser medidas. A partir de una idea general de lo que se pretende investigar, se realizan matizaciones sucesivas a medida que se avanza en el diseño de la investigación para acabar en la formulación de uno o varios objetivos específicos.

La siguiente fase es la elección del diseño básico, es decir, del tipo de estudio más adecuado para responder a la pregunta. Hay que identificar la población diana a la que se desea generalizar los resultados, definir los criterios de selección de los sujetos que participarán en el estudio, estimar el tamaño de la muestra y describir las técnicas que se utilizarán para seleccionarla y para formar los grupos de estudio en caso de que sea necesario.

El protocolo debe especificar las variables y sus definiciones operativas, la técnica de medición que se utilizará y la fuente de información. Habitualmente, una misma variable se puede medir de diferentes maneras y con distintos instrumentos. Es conveniente utilizar definiciones e instrumentos estándar. En cualquier caso, la definición debe ser clara y precisa para que todos los investigadores utilicen criterios homogéneos.

El protocolo ha de describir lo que le ocurrirá a cada persona que participe en el estudio (cómo se incluirá, qué seguimiento se hará, cómo recibirá la intervención, qué mediciones se realizarán, etc.) con el suficiente detalle para que todos los miembros del equipo investigador conozcan lo que debe hacerse en todo momento a lo largo del estudio. Así mismo, debe establecer qué información debe recogerse, quién y de qué forma lo hará, y los mecanismos de control de calidad.

También deben preverse las posibles dificultades que pueden presentarse durante el estudio y especificar las soluciones que se proponen para evitar los sesgos; por ejemplo, técnicas para aumentar la comparabilidad de los grupos, estrategias para disminuir el número de no respuestas o de pérdidas de seguimiento y para recuperarlas si se producen, estrategias que aseguren que la calidad y la homogeneidad de los datos se mantienen a lo largo de todo el estudio, y utilización de técnicas de ciego, si es posible, para evitar posibles interferencias causadas por el investigador o el participante, entre otras soluciones.

En el protocolo también debe constar el plan de análisis de los datos, con la identificación de las variables demográficas que se utilizarán para describir la población estudiada, de los factores pronósticos que se emplearán para valorar la comparabilidad

de los grupos, los subgrupos de sujetos en los que se deseará realizar el análisis, así como las pruebas estadísticas que se aplicarán.

En el protocolo se recogen aspectos relacionados con el método y la organización del estudio, y el calendario previsto. La buena marcha de un estudio depende, en gran medida, de su buena organización, por lo que es importante identificar las tareas administrativas y las personas que las llevarán a cabo. La realización de un estudio suele precisar la colaboración de otras personas, servicios o centros ajenos al propio equipo investigador, por lo que es conveniente contactar con ellos y asegurar su colaboración antes de iniciarlos. El presupuesto ha de ser realista, ya que es un aspecto muy importante para evaluar la viabilidad de un estudio. Es importante asegurar, antes de iniciar el trabajo, que se podrá contar con la financiación necesaria.

El éxito de un trabajo de investigación dependerá de lo bien que se haya diseñado y de si se han identificado los posibles problemas antes de iniciarlos. Además, si el estudio está bien planeado, las fases de recogida y de análisis de los datos serán mucho más fáciles. Por esta razón, antes de iniciar el estudio propiamente dicho, es conveniente realizar una o varias pruebas piloto que pongan de manifiesto los aspectos en los que hay que insistir o que deben mejorarse.

## ELABORACIÓN DEL PROTOCOLO DE ESTUDIO

El protocolo es un documento que se va desarrollando y mejorando a medida que el equipo investigador concreta la idea y el diseño del estudio.

El punto de partida es la identificación del problema y la redacción de una hipótesis genérica. Aunque en esta fase de la investigación esta hipótesis es poco operativa, es conveniente que se ponga por escrito para poder debatirla con otros colegas o miembros del equipo investigador, con el fin de realizar una primera aproximación al desarrollo del objetivo y a la evaluación de la pertinencia y la viabilidad del estudio.

El siguiente paso es desarrollar un esquema general sobre cómo se pretende abordar el problema, e indicar, de forma muy escueta, en 2-4 páginas, los aspectos metodológicos presentados en la cuadro 11.1. Así, por ejemplo, en esta versión, cuando se identifican los criterios de selección de los sujetos, no será necesario definirlos de forma operativa. Igualmente, aunque se deben identificar las principales variables del estudio, no es necesario definirlas con precisión ni indicar detalladamente el método de recogida

de la información. En las versiones sucesivas del protocolo se desarrollarán con más detalle estos aspectos y se matizarán con la incorporación de los elementos organizativos.

Tanto el manual de procedimientos como el protocolo de estudio deben estar siempre disponibles para poder ser consultados ante cualquier duda. Una vez que estos documentos han sido debatidos y consensuados por el equipo investigador, es conveniente que los lean otras personas ajenas al estudio para que juzguen su pertinencia, emitan los comentarios y las sugerencias que crean oportunos, y ayuden a mejorar el proyecto.

En el anexo 6 se presenta, en forma de frases o preguntas, una guía para la elaboración de un protocolo de estudio.

## EQUIPO INVESTIGADOR

Generalmente, en un proyecto de investigación intervienen varias personas que lo impulsan. El equipo ha de ser multidisciplinario, es decir, en él deben estar representados expertos o conocedores de los diferentes aspectos del estudio. Es conveniente que exista un investigador principal que actúe como líder, que debe ser respetado por el resto del equipo, y tener el tiempo y la energía suficientes para llevar adelante el proyecto y mantener el interés del estudio hasta el final.

Entre las funciones del equipo investigador se incluye redactar el protocolo y el manual de procedimientos. De todas formas, es preferible que esta tarea sea asumida por un número reducido de personas, con una de ellas como responsable, y que el protocolo sea consensuado por todo el equipo, con atención a las sugerencias y aportaciones de los miembros. También es conveniente que partes muy especializadas del protocolo, como el análisis estadístico o determinados procedimientos de laboratorio, sean redactados por expertos en la materia.

Desde el inicio, el investigador principal ha de mantener reuniones periódicas con el resto del equipo, con un orden del día pre establecido, en las que se informe de cómo está progresando el diseño, los problemas y las dudas que se plantean, etc. Estas reuniones ayudan a mantener la motivación y el interés de los investigadores. El número y la duración de las reuniones han de ser los imprescindibles, sin ser excesivos.

En estudios multicéntricos, con frecuencia se crean comités para el diseño del proyecto en los que suelen intervenir profesionales de gran prestigio, que difícilmente se harán cargo del desarrollo del estudio en el día a día. Por ello, ha de existir una persona

que asuma este papel y que se ocupe de conocer las convocatorias de ayudas para la financiación, qué centros pueden ser los candidatos para participar, conducir las reuniones de los investigadores, etc.; es decir, que asuma el proyecto como una de sus prioridades y con la estabilidad laboral suficiente como para finalizarlo.

En los estudios multicéntricos, una de las funciones del equipo investigador es identificar y seleccionar los centros y los investigadores que participen en ellos. Es conveniente que esto se haga en fases

tempranas del diseño, como mínimo antes del inicio del estudio, con la finalidad de implicarlos en el diseño o de poder debatir con ellos el protocolo. Una vez se ha contactado con ellos, su aceptación dependerá de varios factores: su interés por el tema, la complejidad del protocolo, la interferencia excesiva en la práctica clínica habitual, la dificultad de la recogida de datos, el número de reuniones, etc. Los incentivos económicos o de otro tipo también pueden decantar el sentido de la respuesta a la solicitud de participación.

## BIBLIOGRAFÍA

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;109:445-51.  
Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman RB, editors

Designing clinical research: an epidemiologic approach. 3rd ed. Baltimore: Lippincott Williams & Wilkins; 2007.  
Knatterud GL. Management and conduct of randomized controlled trials. *Epidemiol Rev*. 2002;24:12-25.

Rebagliato M, Ruiz I, Arranz M. Metodología de investigación en epidemiología. Madrid: Díaz de Santos; 1996.  
Roca J. Cómo y para qué hacer un protocolo. *Med Clin (Barc)*. 1996;106:257-62.

## Objetivo del estudio

Las fases iniciales del desarrollo de un proyecto de investigación son las más importantes. Debe establecerse de forma clara el objetivo que se persigue. Habitualmente se dedica una gran cantidad de esfuerzos a la recogida y el análisis de los datos, pero, si no están bien dirigidos, serán de poca utilidad. Esta etapa requiere una profunda reflexión, ya que muchos aspectos del diseño dependen de lo que se decida al inicio.

### IDENTIFICACIÓN DEL PROBLEMA

Un problema debe entenderse como una incertidumbre sobre algún hecho o fenómeno que el investigador desea resolver realizando mediciones en los sujetos del estudio. Habitualmente, su identificación es fruto de la capacidad del propio profesional para generar ideas y formular interrogantes, y rara vez se produce por pura intuición.

La buena investigación depende, en gran medida, de las buenas ideas. Un investigador debe poseer no sólo talento, sino también capacidad de observación cuidadosa de lo que acontece en su práctica, honestidad y una buena dosis de tenacidad y perseverancia ([cuadro 12.1](#)).

La observación cuidadosa de los pacientes puede detectar sucesos, situaciones o respuestas a tratamientos que parezcan contradecir los conocimientos aceptados, lo que permite sugerir posibles hipótesis de trabajo.

Muchas veces, las preguntas surgen del análisis de estudios previos, propios o de otros autores. El investigador ha de mantener un estado de alerta, realizando una lectura crítica de la literatura, asistiendo a

reuniones científicas y congresos, y colaborando con otros investigadores en un intercambio de ideas y resultados. De esta forma, pueden detectarse lagunas en el conocimiento y generarse interrogantes.

El estado de alerta se complementa con imaginación y creatividad, proponiendo nuevas respuestas a viejas preguntas y manteniendo una actitud escéptica frente a los conocimientos, las creencias y las opiniones que prevalecen en la comunidad científica. La incorporación de nuevas tecnologías diagnósticas o de algunos tratamientos a la práctica clínica se realiza en muchas ocasiones sin haber demostrado previamente su eficacia. Incluso, aunque su uso esté ampliamente difundido, puede considerarse conveniente realizar un estudio que evalúe si un tratamiento es más eficaz o más útil que otras alternativas.

### BÚSQUEDA BIBLIOGRÁFICA

Una vez que se ha centrado el interés en un determinado tema, el siguiente paso es revisar el estado actual del conocimiento sobre él. La investigación no debe entenderse como el intento de responder a una pregunta partiendo de cero, sino que se enmarca en el conjunto de los conocimientos científicos existentes, que son precisamente los que justifican la realización del estudio.

La búsqueda bibliográfica es importante durante todo el proceso de una investigación ([cuadro 12.2](#)). Por un lado, permite saber si la pregunta que se plantea ha sido contestada previamente y, por otro, aprovechar la experiencia previa de otros investigadores para diseñar y ejecutar mejor el estudio.

### Cuadro 12.1 Características de un buen investigador

- Voluntad de investigar
- Talento
- Experiencia
- Capacidad para generar datos
- Honestidad
- Capacidad para obtener recursos
- Flexibilidad
- Perseverancia
- Confianza
- Capacidad para reconocer problemas y desarrollar hipótesis
- Capacidad para congeniar con sus colaboradores
- Competitividad
- Capacidad de redacción

### Cuadro 12.2 Utilidades de la revisión bibliográfica en una investigación

- Fuente de ideas susceptibles de investigación
- Valoración de los conocimientos actuales sobre el tema
- Valoración de la pertinencia y viabilidad del proyecto
- Provisión del marco conceptual para la investigación
- Ayuda en la delimitación del objetivo específico
- Información sobre aspectos concretos del diseño:
  - Estrategias
  - Procedimientos
  - Pautas de seguimiento
  - Criterios de selección
  - Determinación del tamaño de la muestra
  - Definiciones de variables
  - Instrumentos de medición
  - Prevención de problemas
  - Análisis estadístico
- Comparación de los propios resultados con estudios similares
- Contribución a la valoración de la validez externa

Una búsqueda bibliográfica no debe realizarse de forma precipitada ([cuadro 12.3](#)). Es frecuente que el usuario recupere o bien un número escaso o bien un número excesivo de referencias, o que las que recupere no sean adecuadas a sus necesidades. A menudo, esto se debe a una preparación apresurada o imprecisa de la estrategia de búsqueda y a un uso deficiente de la técnica de búsqueda propiamente dicha.

### Cuadro 12.3 Pasos de la elaboración de una estrategia de búsqueda

1. Definir la pregunta e identificar sus componentes
2. Seleccionar la base de datos
3. Seleccionar los términos descriptores y las combinaciones
4. Considerar el uso de restricciones
5. Ejecutar la búsqueda
6. Analizar los resultados
7. Si son inadecuados, revisar la estrategia
8. Si son adecuados, seleccionar los registros de interés y localizar los documentos

Para preparar adecuadamente la estrategia de búsqueda, el primer paso es definir de manera clara y precisa la pregunta de investigación (objetivo) a la que se desea responder. Ello implica identificar con claridad, al menos, los elementos siguientes: problema de salud, enfermedad o condición clínica de interés; características generales de la población; factor de estudio, y en ocasiones también el factor de comparación; y variable de respuesta.

En segundo lugar, deben seleccionarse las bases de datos bibliográficas en las que se realizará la búsqueda. Dada su accesibilidad y cobertura, probablemente MEDLINE sea la alternativa principal, aunque en ocasiones otras pueden ser más adecuadas. Por ejemplo, la Cochrane Library puede ser más útil para localizar revisiones sistemáticas. Dado que los principales repertorios bibliográficos tienen una cobertura distinta, si se quiere realizar una búsqueda exhaustiva se deben intentar recuperar los artículos a partir de más de una fuente, aunque ello produzca un elevado porcentaje de artículos duplicados.

En tercer lugar, deben trasladarse los elementos de la pregunta de investigación al lenguaje de la base de datos. El conocimiento del tesoro utilizado por la base de datos que se consulta es necesario para recuperar todos los documentos existentes en ella sobre el tema de interés.

La selección de los términos descriptores empieza identificando los que mejor representan la materia de interés, así como términos afines y técnicas y procedimientos de especial relevancia, y estableciendo sus relaciones mutuas.

Los términos descriptores pueden combinarse según los posibles operadores de relación o conexión. Habitualmente se utilizan los operadores lógicos o «booleanos» AND (los dos términos que conecta deben estar presentes en el documento), OR (es suficiente que esté presente uno de los términos) y NOT (excluye los documentos en los que esté presente el término al que precede). Estos operadores deben

emplearse con cuidado, ya que de ellos depende, en gran medida, la eficacia de la búsqueda.

Los programas de consulta de la base de datos permiten ejecutar la búsqueda aplicando determinadas restricciones, como el campo en que deben aparecer los términos, el idioma o la fecha de publicación o, por ejemplo, el tipo de publicación (ensayo clínico, revisión sistemática, carta, etc.).

Si los artículos identificados no son los adecuados, debe revisarse la estrategia, porque es probable que los términos no se hayan seleccionado y/o combinado adecuadamente.

Si el número de artículos identificados es reducido, debe comprobarse que los términos se hayan escrito correctamente y valorar la utilización de descriptores más genéricos o de otros términos relacionados no incluidos. También puede ser útil emplear lenguaje natural o incluso suprimir algunas de las restricciones utilizadas.

Si el número de artículos es excesivo, debe plantearse el uso de descriptores más específicos y/o su combinación con otros términos (mediante el operador AND). También puede revisarse el uso de limitadores más restrictivos y de filtros metodológicos.

## DELIMITACIÓN DEL OBJETIVO

La definición precisa del objetivo es el eje en torno al cual se diseña el estudio. Si no es suficientemente clara, será difícil tomar decisiones sobre el tipo de estudio más apropiado, la población que se va a incluir o las variables que deberán medirse.

La elaboración del modelo teórico en que se enmarca la investigación ayudará a identificar las variables. Este proceso suele requerir debates con otros profesionales, valorar sus sugerencias y hacer una lectura de la bibliografía. Inicialmente, se formula el objetivo en términos genéricos y se establece una pregunta amplia no susceptible de investigación específica. Sirve para poco más que para centrar el tema de interés. Después, se acota progresivamente hasta formular una pregunta precisa que permita evaluar una hipótesis concreta. La formulación explícita de este objetivo específico determina la planificación subsiguiente (fig. 12.1).

**Ejemplo 12.1.** Se desea estudiar la utilidad de la educación sanitaria en el tratamiento de los pacientes diabéticos. Los investigadores asumen el siguiente modelo teórico:

Educación sanitaria → Aumento de conocimientos → Cambio de actitudes → Aumento del cumplimiento de las recomendaciones (dieta, autocontroles, etc.) → Mejoría del control metabólico →



Figura 12.1 Delimitación del objetivo del estudio.

Disminución de las complicaciones → Aumento de la esperanza y la calidad de vida.

A partir de este marco general, los investigadores deberán decidir el aspecto que tiene más interés o es más factible estudiar. Dado que el estudio de las complicaciones requiere un seguimiento prolongado, deciden evaluar el efecto de la educación sanitaria sobre el control metabólico de la diabetes, asumiendo que un mejor control implica una disminución de las complicaciones y un aumento de la esperanza y la calidad de vida. A pesar de esta asunción, las conclusiones del estudio sólo pueden ceñirse a la influencia de la educación sanitaria sobre el control metabólico. La elaboración de este modelo teórico tiene la ventaja de que identifica algunas de las variables que será importante medir en el estudio, como el tipo de conocimientos o el cumplimiento de las recomendaciones.

## PERTINENCIA Y VIABILIDAD

En las fases iniciales deben valorarse la pertinencia y la viabilidad del proyecto, de forma que pueda justificarse su realización (cuadro 12.4).

Debe asegurarse que la pregunta no ha sido contestada previamente o lo ha sido de forma contradictoria. No deben realizarse estudios sólo en función de los temas contemplados en las jornadas y los congresos científicos, sino que deben orientarse a la resolución de problemas frecuentes o importantes. Es conveniente que sean coherentes con las líneas de investigación, lo que permitirá aprovechar los

**Cuadro 12.4 Aspectos a valorar en la justificación de un proyecto de investigación****Pertinencia**

- La pregunta planteada no ha sido contestada, lo ha sido de forma contradictoria, o se requiere adecuarla a la propia práctica
- Importancia del problema:
  - Frecuencia, gravedad, interés social, etc.
  - Beneficios que pueden derivarse del estudio
  - Aplicaciones prácticas que pueden derivarse de los resultados
  - Importancia por su valor teórico
- La pregunta y el diseño del estudio se ajustan a los principios éticos de investigación
- Los resultados esperables compensan los recursos utilizados

**Viabilidad**

- Variables que sea factible medir
- Tiempo hasta la aparición del resultado
- Oportunidad
- Disponibilidad de sujetos (tanto en número como en el tiempo necesario para su inclusión)
- Colaboración con otros profesionales, servicios o centros
- Instalaciones y equipamiento
- Recursos económicos (financiación)
- Experiencia del equipo investigador
- Consideraciones éticas
- Interés del propio equipo investigador

recursos con una mayor eficiencia y profundizar en el conocimiento del problema.

El propósito de la mayoría de las investigaciones clínicas y epidemiológicas es servir de base para la toma de decisiones, ya sea inmediata o a largo plazo. La potencial utilidad clínica y la aplicabilidad práctica de los resultados deben tenerse muy en cuenta.

En la evaluación de la pertinencia, los investigadores deben asegurarse de que tanto la pregunta como el tratamiento que se plantea estén de acuerdo con los principios éticos de la investigación en seres humanos.

La valoración de la viabilidad es un proceso más dinámico, que requiere considerar aspectos que serán afrontados en fases más avanzadas de la planificación. En este momento, debe concebirse la pregunta como una formulación provisional que puede ser replanteada o modificada en función de las dificultades que vayan apareciendo durante el diseño y en la prueba piloto.

Debe valorarse la capacidad del investigador para medir las variables de interés, el número de sujetos necesarios, el tiempo preciso y los recursos humanos, de equipamiento y económicos requeridos. La insuficiencia de estos medios o la incapacidad para reunir el número de sujetos necesario en un tiempo determinado hace que, con frecuencia, se requiera la colaboración de otros centros o servicios. Debe asegurarse esta colaboración y coordinación antes de iniciar el estudio.

La oportunidad es un criterio que facilitará la realización del estudio. Esta oportunidad puede entenderse desde diferentes vertientes; por ejemplo, por la disponibilidad de sujetos suficientes al existir una epidemia de la enfermedad, porque el objetivo del estudio corresponda a alguna de las líneas prioritarias de las agencias financieras, lo que permitirá obtener los recursos necesarios con más facilidad, o bien porque el tema de estudio sea de actualidad o tenga una gran repercusión social.

**OBJETIVO ESPECÍFICO**

El objetivo específico es la pregunta principal que se quiere contestar, acorde con el propósito del estudio, formulada con precisión y sin ambigüedades y en términos mensurables, es decir, realistas y operativos. Debe concebirse como una hipótesis contrastable, al menos dentro de las limitaciones del método disponible.

El objetivo específico lleva implícita la respuesta que propone el investigador. De hecho, cuando se plantea un estudio, el investigador parte de una hipótesis previa. Por ejemplo, si desea evaluar la eficacia de un fármaco es porque cree que es más eficaz que el placebo o que otros tratamientos. Debe evitarse que esta predisposición pueda influir sobre las observaciones y mantener la objetividad en todo momento durante el estudio.

Un estudio suele estar diseñado para responder a una única pregunta. Si existe más de una, la principal debe ser aquella que el investigador esté más interesado en responder y que pueda ser bien contestada, y el resto puede considerarse como preguntas secundarias.

Una buena pregunta principal debe cumplir las siguientes condiciones: ser única, relevante, simple, consistente, novedosa, viable, contrastable, ética, definida a priori y formulada explícitamente.

**Formulación del objetivo específico**

La formulación del objetivo específico de un estudio descriptivo debe incluir el fenómeno que se

desea describir (p. ej., la prevalencia, la incidencia, la supervivencia, los síntomas de presentación, la evolución natural, etc.), el problema de salud y la población objeto de estudio.

**Ejemplo 12.2.** Objetivo del estudio: estimar la prevalencia de hipertensión arterial en una zona básica de salud. Esta formulación del objetivo informa que se desea estudiar la frecuencia de hipertensión arterial mediante la prevalencia, y que la población objeto de estudio son los residentes en una zona básica de salud.

**Ejemplo 12.3.** Objetivo del estudio: describir la supervivencia de los pacientes con enfermedad pulmonar obstructiva crónica que reciben oxigenoterapia domiciliaria. El fenómeno que se estudia es la supervivencia. La población no se define geográficamente, sino que corresponde a los sujetos con un determinado problema de salud y que reciben oxigenoterapia domiciliaria.

**Ejemplo 12.4.** Objetivo del estudio: evaluar si la concentración sérica del antígeno prostático específico permite distinguir a los pacientes mayores de 50 años con cáncer de próstata de los que no lo padecen, utilizando como método de referencia el tacto rectal y el estudio anatomo-patológico del material obtenido por punción transuretral. Obsérvese que se trata de un estudio que evalúa una prueba diagnóstica (concentración sérica del antígeno prostático específico). En este tipo de estudios, en el objetivo debe constar la cuestión clínica precisa que se pretende resolver con la aplicación de la prueba diagnóstica, la naturaleza de la población estudiada y el método utilizado como criterio de referencia.

En un estudio analítico, la formulación del objetivo específico requiere la identificación de los siguientes elementos: el *factor de estudio*, que corresponde a la exposición o intervención de interés; la *variable de respuesta*, o criterio de evaluación, con la que se pretende medir el efecto o la asociación con dicho factor, y la *población*, o conjunto de sujetos a los que se desea aplicar los resultados del estudio. En ocasiones, puede ser conveniente incluir en la formulación el factor o los factores que se utilizarán como comparación. En esta fase del diseño, basta con una definición genérica de estos elementos. A medida que se avance en la elaboración del protocolo se irán definiendo con más detalle sus características.

**Ejemplo 12.5.** Objetivo del estudio: evaluar si la educación sanitaria grupal produce mejores resultados que la individual en el control metabólico de los

pacientes diabéticos tipo 2. El factor de estudio es la educación sanitaria grupal, cuyo efecto se comparará con la individual, que actuará como referencia. La variable de respuesta es el control metabólico, mientras que la población corresponde a los sujetos con diabetes mellitus tipo 2.

**Ejemplo 12.6.** Objetivo del estudio: determinar si el consumo de tabaco aumenta la incidencia de cardiopatía isquémica en sujetos hipertensos. El factor de estudio es el consumo de tabaco, la variable de respuesta es el desarrollo de cardiopatía isquémica y la población son los sujetos hipertensos.

Es frecuente que el objetivo de un estudio esté mal definido. A veces, el problema radica simplemente en su deficiente redacción. Ello no implica que el estudio esté mal diseñado, pero un objetivo mal escrito no ayuda a su comprensión.

**Ejemplo 12.7.** En un artículo se definía el objetivo del estudio del siguiente modo: «El propósito de nuestro estudio radica en el conocimiento de los tumores del sistema nervioso central en el marco de una comunidad autónoma». Este objetivo se expresa en términos demasiado vagos. Debería especificar si se refiere, por ejemplo, a su epidemiología, las características clínicas, el pronóstico o los factores de riesgo.

**Ejemplo 12.8.** Supongamos un estudio cuyo objetivo se define como «determinar los efectos de la vacuna frente al virus de la hepatitis A». Definido de esta forma, no se puede saber qué pretenden los autores. Puede tratarse tanto de un estudio descriptivo de los efectos secundarios de la vacuna como de un estudio experimental que evalúe su eficacia. Además, tampoco da ninguna orientación sobre el tipo de sujetos en quienes se realiza el estudio.

En otras ocasiones, se incluyen como objetivos de investigación cuestiones que no lo son. En realidad, un objetivo específico de investigación no es más que una pregunta concreta a la que se pretende responder mediante la realización del estudio. No debe confundirse con los objetivos de un programa sanitario, que consisten en la modificación de un estado de salud en un colectivo.

**Ejemplo 12.9.** Supongamos un estudio que define sus objetivos como: a) conocer el tipo de reacciones adversas a medicamentos notificadas por los facultativos de atención primaria; b) identificar su posible infranotificación, y c) proponer medidas para intentar mejorar dicha notificación. El tercer

objetivo corresponde más a recomendaciones que realizan los autores que a un objetivo de investigación, ya que no implica ninguna pregunta a la que el estudio pueda dar respuesta.

En otras ocasiones, el objetivo no existe. El investigador se limita a recoger datos y analizarlos sin una estrategia previa. Este solo hecho compromete la validez de cualquier estudio.

### Variable de respuesta

De los tres elementos que componen el objetivo específico de un estudio analítico, el que plantea más problemas es la variable de respuesta. Se puede considerar adecuada si mide el fenómeno que se desea estudiar y puede ser valorada con un instrumento adecuado y en todos los sujetos del estudio (cuadro 12.5).

Existen múltiples situaciones en las que la elección de la variable más adecuada puede resultar una tarea difícil (cuadro 12.6). Estas dificultades pueden estar relacionadas tanto con la propia variable como con las limitaciones inherentes al proceso de su medición.

A menudo interesa estudiar fenómenos que incluyen un gran componente de subjetividad, como puede ser la evolución del dolor, la mejoría de síntomas clínicos o la calidad de vida. A veces es difícil definir una variable que mida estos conceptos con las suficientes validez y precisión.

**Ejemplo 12.10.** Supongamos un estudio sobre la eficacia de un tratamiento analgésico. Dado que no existe un método objetivo para determinar la intensidad del dolor, los autores se ven obligados a utilizar una escala visual analógica, que consiste en

### Cuadro 12.5 Características de una buena variable de respuesta

- Estar definida con precisión y antes de iniciar el estudio
- Ser apropiada a la pregunta que se desea responder
- Medir lo que se quiere medir
- Ser suficientemente sensible para detectar el efecto de interés
- Poderse medir tan detallada como sea posible
- Poderse medir con un método fiable, preciso y reproducible
- Poderse medir en todos los sujetos y de la misma manera
- Ser única en la medida de lo posible. Si hay varias, es preferible utilizar la más relevante y fiable

### Cuadro 12.6 Dificultades en la elección de la variable de respuesta

#### Relacionadas con la variable

- Fenómenos que no pueden medirse objetivamente
- Utilización de variables aproximadas
- Utilización de variables intermedias
- Existencia de más de una variable
- Medida parcial del fenómeno
- Uso de escalas combinadas
- Uso de variables que no miden el fenómeno de interés

#### Relacionadas con la medición

- Variabilidad inter e intraobservador
- Variabilidad inter e intraindividual
- Precisión y exactitud del instrumento de medida

una representación gráfica en forma de línea recta cuyos extremos definen los valores límite de la intensidad del dolor. Los pacientes marcan en la escala el punto que, a su juicio, representa la intensidad de su síntoma.

En otras ocasiones, la mejor variable de respuesta no está al alcance de los investigadores y se utiliza una variable aproximada.

**Ejemplo 12.11.** Consideremos un estudio en el que se evalúa el grado de cumplimiento de un tratamiento tras una intervención educativa. Supongamos que se pueden medir los valores del fármaco o sus metabolitos en orina. Aunque ésta sería la variable de respuesta de elección, si la prueba no está disponible, es muy cara o incómoda para los pacientes, los autores pueden optar por la utilización de una variable de respuesta aproximada, como la información proporcionada por el propio paciente sobre el cumplimiento del tratamiento.

En ocasiones se recurre al uso de una variable intermedia o subrogada en sustitución de la variable primaria de interés, porque se cree que está correlacionada con ella, refleja la totalidad o gran parte del efecto, y puede obtenerse con un coste inferior, en un período más breve, o en circunstancias en las que aquella no podría estudiarse. Los cambios observados en la variable intermedia o subrogada se supone que reflejan los que se producirían en la variable primaria. El uso de estas variables lleva implícita la asunción de la existencia de una relación directa con la respuesta final. Así, el uso de la

reducción de las cifras de colesterolemia total tras la administración de un fármaco se hace bajo la asunción de que comporta la disminución de la morbilidad cardiovascular. Si esta asunción no es válida, la variable intermedia o subrogada tampoco lo es, e incluso, si la asunción es válida, estrictamente la conclusión sólo podrá aplicarse a la variable intermedia.

**Ejemplo 12.12.** Supongamos un estudio que evalúa la eficacia de una técnica de educación sanitaria en sujetos hipertensos. Los investigadores utilizan como variable de respuesta los conocimientos de los pacientes después de la educación y asumen que un mejor conocimiento comporta un mejor control de la hipertensión arterial y, por ello, una disminución de la morbilidad cardiovascular. Esta asunción puede ser cuestionable. De hecho, cuanto más alejada esté una variable intermedia o subrogada del resultado final, más probable es que no sea válida.

Puede ocurrir que existan diferentes variables que midan aspectos distintos del fenómeno de interés. Estas situaciones pueden plantear problemas, ya que si se obtienen resultados contradictorios o inconsistentes para cada una de ellas, la interpretación resulta difícil. La mejor solución es haber definido a priori una variable principal. Esto no quiere decir que no puedan estudiarse todas las variables simultáneamente, sino que deberá decidirse cuál de ellas es la más importante para el investigador, y en la que se basará la conclusión del estudio.

**Ejemplo 12.13.** En un estudio que evaluaba la eficacia de un tratamiento preventivo de las crisis convulsivas febriles en la infancia, los investigadores utilizaron como variable de respuesta la frecuencia de las crisis. Ello no significa que no midieran otros aspectos como la gravedad de las crisis, sino que la conclusión sobre la eficacia del tratamiento se basó en la disminución de su frecuencia.

Otro problema que puede presentarse es el uso de variables que sólo midan parcialmente el fenómeno de interés.

**Ejemplo 12.14.** En un estudio se comparaba el efecto de dos estrategias de atención sanitaria sobre el estado de salud de un grupo de pacientes ancianos. Como variable de respuesta se utilizó una escala que valoraba su estado funcional. Sin embargo, esta variable sólo mide parcialmente las necesidades de salud de los ancianos y no tiene en cuenta, por ejemplo, su estado cognitivo. Como siempre, las

conclusiones del estudio deberán limitarse a la variable de respuesta, que en este estudio es el estado funcional de los sujetos.

Cuando no existe una variable única que por sí sola exprese toda la complejidad del fenómeno que se desea medir, se pueden utilizar escalas o índices compuestos en los que se agrupan variables de naturaleza diferente (p. ej., escalas para medir la autonomía para realizar las actividades de la vida diaria, la calidad de vida, la salud física o mental, etc.). Es importante que este tipo de escalas hayan sido validadas en el medio en que se va a realizar el estudio.

En otras ocasiones se usa una variable de respuesta principal, que es la suma o la combinación de diferentes variables que se supone que reflejan el fenómeno o los fenómenos de interés. Por ejemplo, un estudio sobre prevención cardiovascular podría utilizar como variable principal de respuesta el número de episodios cardiovasculares, definido como la suma de episodios de infarto agudo de miocardio (IAM) mortal, IAM no mortal, muerte súbita, accidente cerebrovascular (ACV) mortal y ACV no mortal. Una variable combinada de esta naturaleza permite aumentar el poder estadístico del estudio, pero hay que ser muy prudente a la hora de interpretar los resultados y es recomendable analizar cada una de las variables de forma separada.

Otro problema distinto es la utilización de una variable de respuesta que no mida el fenómeno de interés, aunque existan otras variables válidas, precisas y factibles de medir.

**Ejemplo 12.15.** En un estudio se pretende evaluar la eficacia de una intervención educativa sobre el control metabólico de los pacientes diabéticos y, en lugar de utilizar la hemoglobina glucosilada como variable de respuesta, se escoge una determinación aislada de glucemia, que es menos válida y precisa.

## PREGUNTAS SECUNDARIAS

La pregunta principal puede complementarse con otras preguntas secundarias. Algunas de ellas harán referencia a subgrupos de sujetos de especial interés y otras requerirán variables de respuesta diferentes a la principal, ya sea complementándola o midiendo efectos diferentes. Deben fijarse a priori para evitar el riesgo de no recoger toda la información necesaria y limitar su número, ya que su estudio supone incrementar la complejidad del diseño y su realización. De todas formas, dado que el estudio ha sido

diseñado para responder a la pregunta principal, la respuesta a las preguntas secundarias será menos «sólida».

En su definición es de gran ayuda haber establecido el marco teórico. En el ejemplo 12.1 sobre la educación sanitaria en los pacientes diabéticos, el interés se centraba en el control metabólico, pero sería interesante estudiar otros aspectos del modelo. Por ejemplo, los conocimientos, ya que puede ser útil evaluar si, en el caso de no poderse demostrar la

eficacia de la técnica esto se debe a que la educación sanitaria no ha sido eficaz, o bien a que los pacientes no han comprendido la información proporcionada. Además, pueden estudiarse otros aspectos colaterales, como las características de los pacientes que no acuden a las sesiones grupales, si el incremento del grado de conocimientos se acompaña de un mejor control metabólico de la enfermedad, etc., de forma que pueda obtenerse el máximo rendimiento de la información recogida.

## BIBLIOGRAFÍA

- Bakke OM, Carné X, García Alonso F. *Ensayos clínicos con medicamentos: fundamentos básicos, metodología y práctica*. Barcelona: Doyma; 1994. p. 116-26.
- Berger VW. Improving the information content of categorical clinical trial endpoints. *Control Clin Trials*. 2002;23:502-14.
- Bethel MA, Holman R, Haffner SM, Califf RM, Huntsman-Labed A, Hua TA, et al. Determining the most appropriate components for a composite clinical trial outcome. *Am Heart J*. 2008;156(4):633-40.
- Cooper R, Kaanders JH. Biological surrogate end-points in cancer trials: potential uses, benefits and pitfalls. *Eur J Cancer*. 2005;41(9):1261-6.
- Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ*. 2010;341:c3920.
- Freeman N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA*. 2003;289(19):2554-9.
- Grimes DA, Schulz KF. Surrogate end points in clinical research: hazardous to your health. *Obstet Gynecol*. 2005;105(5 Pt 1):1114-8.
- Hopewell S, Clarke M, Lefebvre C, Scherer R. Handsearching versus electronic searching to identify reports of randomized trials. *Cochrane Database Syst Rev*. 2007;(2):MR000001.
- Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman RB, editors. *Designing clinical research: an epidemiologic approach*. 3rd ed. Baltimore: Lippincott Williams & Wilkins; 2007.
- Huque MF, Alesh M, Bhore R. Addressing multiplicity issues of a composite endpoint and its components in clinical trials. *J Biopharm Stat*. 2011;21(4):610-34.
- Kassai B, Shah NR, Leizorovicza A, Cucherat M, Gueyffier F, Boissel JP. The true treatment benefit is unpredictable in clinical trials using surrogate outcome measured with diagnostic tests. *J Clin Epidemiol*. 2005;58(10):1042-51.
- Parkhill AF, Clavisi O, Patuwave L, Chau M, Turner T, Bragge P, et al. Searches for evidence mapping: effective, shorter, cheaper. *J Med Libr Assoc*. 2011;99(2):157-60.
- Staunton M. Evidence-based radiology: steps 1 and 2--asking answerable questions and searching for evidence. *Radiology*. 2007;242(1):23-31.
- Tassinari D. Surrogate end points of quality of life assessment: have we really found what we are looking for? *Health Qual Life Outcomes*. 2003;1:71.
- Wittkop L, Smith C, Fox Z, Sabin C, Richert L, Aboulker JP, et al. Methodological issues in the use of composite endpoints in clinical trials: examples from the HIV field. *Clin Trials*. 2010;7(1):19-35.

## Variable de respuesta

La definición del objetivo específico del estudio incluye la identificación, en términos genéricos, de la variable principal que se utilizará para evaluar y cuantificar los efectos del factor de estudio y compararlos con los del grupo de referencia para responder a la pregunta de investigación. La elección de esta variable es un aspecto clave en el diseño de un estudio, ya que se utilizará para interpretar la relevancia clínica de los resultados.

La variable principal debería ser aquella capaz de proporcionar la evidencia más convincente y clínicamente relevante, directamente relacionada con el objetivo del estudio. No debe seleccionarse tan sólo porque sea fácil de medir o porque se espera que pueda mostrar cambios o diferencias en un corto período de tiempo, sino que es importante elegir la más adecuada, es decir, la que mida los verdaderos resultados de importancia e interés para los pacientes. Debería haber una única variable principal, que pueda medirse de forma válida y fiable, y su selección debería reflejar las normas y estándares aceptados en el ámbito de investigación específica.

**Ejemplo 13.1.** El Committee for Medicinal Products for Human Use (CHMP) ha elaborado unas recomendaciones para la evaluación de los nuevos medicamentos en el tratamiento de la osteoporosis primaria (CHMP, 2005), entre las que se incluye la de que todos los desenlaces para evaluar la eficacia en los ensayos clínicos deben ser definidos antes del inicio del ensayo y estar adecuadamente incluidos en el protocolo.

Así, por ejemplo, en relación con las fracturas, establece que la variable principal debería estar basada en la aparición de nuevas fracturas, y no en el empeoramiento de las previas (aunque la existencia de fracturas y deformidades previas también debería

registrarse), y que en los ensayos confirmatorios deberían estudiarse separadamente las fracturas vertebrales y las no vertebrales. Dado que la unidad de análisis debe ser el paciente, y no la fractura, la variable principal debería definirse como la incidencia de pacientes con nuevas fracturas. Así mismo, recomienda la realización de radiografías anuales para detectar la aparición de fracturas y deformidades, así como la práctica de exploraciones radiológicas adicionales para identificar fracturas vertebrales sintomáticas; por ello, resulta imprescindible la estandarización de los procedimientos para minimizar la variabilidad.

Las recomendaciones también establecen que la densidad mineral ósea puede ser la variable principal en los estudios exploratorios, pero que no es una variable subrogada adecuada de la reducción de fracturas. También establecen que la estatura puede utilizarse como una variable secundaria, y que los marcadores bioquímicos de recambio óseo pueden ser útiles para evaluar el mecanismo de acción del fármaco, pero no pueden considerarse variables principales adecuadas en los ensayos terapéuticos confirmatorios.

**Ejemplo 13.2.** OMERACT es el acrónimo de una iniciativa internacional dirigida a mejorar la medida de los resultados en la investigación clínica del área de la reumatología (Tugwell et al, 2007). Se inició en 1992 para establecer un consenso en relación con la artritis reumatoide en tres aspectos: el número mínimo de medidas de resultado que deberían incluirse en los ensayos clínicos, la magnitud de la diferencia que podría considerarse clínicamente importante y el grado en que clínicos e investigadores consideraban que los índices compuestos podían ser útiles. Para ser avaladas por dicha iniciativa, las medidas deben pasar un triple filtro: validez (¿mide lo que pretende medir? ¿El resultado es relevante y no sesgado?),

discriminación (¿discrimina entre situaciones de interés, tanto con finalidades de clasificación o pronóstico como de detección de cambios?) y factibilidad (¿puede aplicarse fácilmente, dadas las restricciones de tiempo, económicas y de interpretabilidad?). Desde entonces, han elaborado consensos también en otros ámbitos, como la artrosis, la osteoporosis o la artritis psoriásica, por ejemplo.

Existen iniciativas similares en otras áreas, por ejemplo las recomendaciones IMMPACT en el ámbito de la investigación clínica en dolor (Dworkin et al, 2005).

La variable principal debe definirse cuidadosamente. No sería adecuado especificar simplemente qué es la mortalidad, sin indicar si se utilizará la proporción de sujetos que siguen vivos en un momento determinado o la distribución del tiempo de supervivencia en un período especificado, por ejemplo. O bien, en el caso de que se trate de sucesos recurrentes como el infarto agudo de miocardio (IAM), debería especificarse si se refiere a cualquier aparición del mismo en un paciente durante un período determinado, al tiempo transcurrido hasta el primer episodio en un sujeto, al número de sucesos por unidades-tiempo de observación, etc.

Aunque a menudo la selección de una única variable puede ser difícil, es importante evitar los problemas relacionados con la multiplicidad, por lo que el protocolo deberá ser muy claro al respecto, identificando y definiendo con claridad cuál es la variable principal que se utilizará en el análisis estadístico y en el cálculo del tamaño de la muestra, y cuáles las secundarias.

**Ejemplo 13.3.** D'Haens et al (2007) revisaron las diferentes variables utilizadas para evaluar la eficacia de nuevos fármacos en el tratamiento de la colitis ulcerosa, y observaron que su multiplicidad dificultaba la comparación de los resultados de los diferentes estudios, así como su agregación en metaanálisis. Cooney et al (2007) analizaron las características de los diferentes índices utilizados, y detectaron que existía una amplia heterogeneidad y que diferían en el uso de medidas objetivas (frecuencia de deposiciones, temperatura, frecuencia cardíaca, etc.) y subjetivas (valoración clínica global, bienestar general, etc.), así como en la utilización de la sigmoidoscopia. Encuentran que los diversos índices no estaban validados y que no existía una definición de remisión de la enfermedad internacionalmente aceptada ni consenso en los sistemas de puntuación de los hallazgos clínicos, endoscópicos e histológicos. Los autores llegaron a la conclusión de que era necesaria una estandarización de las medidas de la respuesta.

**Ejemplo 13.4.** Los ensayos clínicos con fármacos quimioterápicos en casos de cáncer avanzado pueden utilizar diferentes criterios para evaluar la respuesta: relacionados con el propio cáncer, como el tiempo de supervivencia, el tiempo hasta la respuesta tumoral o la duración de la misma; relacionados con el riesgo de infección, como la toxicidad hematológica; o relacionados con la calidad de vida del paciente, como cambios en el estado funcional o la aparición de efectos adversos, por ejemplo.

Aunque todas las variables anteriores proporcionan algún tipo de información útil sobre el efecto de un tratamiento, los ensayos clínicos de fase III suelen utilizar el tiempo de supervivencia como variable principal de respuesta. Sin embargo, los ensayos de fase I, dado que se centran en establecer la dosis más adecuada, se interesan más en la toxicidad, mientras que los de fase II se centran más en la respuesta tumoral.

A menudo interesa estudiar fenómenos que incluyen un gran componente de subjetividad, como puede ser la evolución del dolor, la mejoría de síntomas clínicos o la calidad de vida. A veces es difícil definir una variable que mida estos conceptos con la suficiente validez y precisión. Cuando la mejor variable de respuesta no está al alcance de los investigadores, puede utilizarse una variable aproximada.

También puede ocurrir que existan diferentes variables que midan aspectos distintos del efecto de interés, y que los investigadores deseen utilizar más de una variable principal, cada una de las cuales (o un conjunto de ellas) podría ser suficiente para evaluar el rango de efectos de interés. Estas situaciones pueden plantear problemas, ya que, si se obtienen resultados contradictorios o inconsistentes para cada una de ellas, la interpretación resulta difícil. Por ello, debe haberse definido muy claramente a priori la forma de interpretar las evidencias, indicando si es necesario que exista impacto sobre una sola, sobre un número determinado o sobre todas ellas para considerar que ha existido una respuesta positiva, así como los problemas potenciales relacionados con las comparaciones múltiples o la existencia de correlación entre las variables.

Cuando no existe una variable única que, por sí sola, exprese toda la complejidad del fenómeno, se pueden utilizar escalas o índices compuestos, en los que se agrupan variables de naturaleza diferente (p. ej., escalas o cuestionarios para medir la autonomía para realizar las actividades de la vida diaria, la calidad de vida, la salud física o mental, etc.). Es importante que este tipo de escalas hayan sido validadas en el medio en que se va a realizar el estudio. En otras ocasiones se usa una variable de respuesta principal que es la suma de diferentes

variables que se supone que reflejan el fenómeno o los fenómenos de interés (variable combinada).

Por otro lado, las variables de respuesta pueden ser de varios tipos (**tabla 13.1**). Aunque lo que realmente interesa medir es el impacto sobre la enfermedad y el paciente, en muchas ocasiones se recurre al uso de variables intermedias o subrogadas, que habitualmente corresponden a medidas de laboratorio o signos físicos (como la colesterolemia o las cifras de presión arterial), que se utilizan como sustitutos de un resultado clínicamente relevante. Existe una importante discusión sobre la adecuación del uso de este tipo de variables, ya que muchos de los resultados clínicos de interés tienen una etiología multifactorial y pueden existir diversos factores de confusión que interfieran en su relación con las variables subrogadas, de manera que existen múltiples situaciones en las que éstas no reflejen adecuadamente los efectos sobre el resultado clínico de interés.

Las variables clínicas se definen sobre la base de la enfermedad en estudio, como la supervivencia en el caso del cáncer o las fracturas en el de la osteoporosis, por ejemplo. Cuando se refieren a síntomas o medidas funcionales (frecuencia de la angina, tolerancia al ejercicio, etc.) que no son los resultados finales de interés, pueden considerarse variables intermedias.

Las variables relevantes para los pacientes miden resultados que son importantes para los sujetos, como la calidad de vida o la recuperación de una capacidad funcional determinada. A veces también pueden considerarse en esta categoría variables clínicas o incluso subrogadas, siempre que sean buenas predictoras en un sentido causal de estos resultados. Muchos estudios se centran en evaluar el efecto

sobre una variable principal y un número limitado de resultados secundarios para mostrar los beneficios potenciales de un tratamiento, pero a menudo no miden sus efectos sobre otras variables importantes o incluso algunos efectos secundarios y adversos. La medición de variables que incluyan la perspectiva del paciente será importante para establecer un adecuado balance entre beneficios y riesgos.

## EVALUACIÓN DE LA RESPUESTA

Los aspectos relacionados con la medición e interpretación clínica de la variable de respuesta deben ser justificados en el protocolo. Así, su definición debe ir acompañada de la identificación del instrumento o procedimiento que se utilizará para medirla, ya que en buena parte los resultados del estudio dependen de que se disponga de datos fiables de cada paciente.

La respuesta de un paciente ha de medirse de una manera objetiva, precisa y consistente, de manera que el estudio en su conjunto proporcione una evaluación adecuada del efecto que se estudia. Por ello, los procedimientos de medida y registro de los datos han de estar muy bien establecidos en el protocolo. Dado que es necesaria una evaluación más rigurosa de lo que sería habitual en la práctica clínica, los registros de rutina no suelen ser adecuados para una correcta evaluación, ya que suelen ser vagos, inconsistentes y subjetivos.

Idealmente, todas las medidas realizadas sobre un paciente deben ser precisas y reproducibles, y no deberían depender del observador, por lo que deben

**Tabla 13.1 Ejemplos de variables subrogadas, clínicas y relevantes para los pacientes**

ENFERMEDAD	RESULTADO		
	Subrogado	Clínico	Relevante para el paciente
Cardiovascular	Presión arterial	<ul style="list-style-type: none"> <li>• ACV, IAM</li> <li>• Supervivencia</li> </ul>	<ul style="list-style-type: none"> <li>• Resultados clínicos</li> <li>• Calidad de vida</li> </ul>
VIH/sida	Recuento de CD4	<ul style="list-style-type: none"> <li>• Eventos relacionados con sida</li> <li>• Supervivencia</li> </ul>	<ul style="list-style-type: none"> <li>• Supervivencia</li> <li>• Calidad de vida</li> <li>• Efectos adversos</li> </ul>
Fractura	Densidad mineral ósea	<ul style="list-style-type: none"> <li>• Fractura ósea</li> </ul>	<ul style="list-style-type: none"> <li>• Fractura sintomática</li> <li>• Calidad de vida</li> <li>• Estado funcional</li> </ul>
Coronaria	Colesterolemia	<ul style="list-style-type: none"> <li>• IAM</li> <li>• Supervivencia</li> </ul>	<ul style="list-style-type: none"> <li>• IAM</li> <li>• Supervivencia</li> <li>• Síntomas (angina)</li> <li>• Calidad de vida</li> <li>• Dolor</li> </ul>

adoptarse todas las estrategias posibles para evitar al máximo cualquier fuente de variación: estandarización de procedimientos, homogeneización de criterios, entrenamiento de los observadores, calibración de los dispositivos, laboratorio centralizado, etc. En los capítulos 18 y 19 se abordan con detalle los aspectos relacionados con la medición de variables.

**Ejemplo 13.5.** En el ámbito de la iniciativa IMMPACT, para estandarizar las medidas de resultado en los ensayos clínicos sobre dolor crónico, se recomienda que se consideren seis dominios al diseñar los estudios: dolor, funcionalidad física, funcionalidad emocional, percepción del paciente sobre la mejoría y la satisfacción con el tratamiento, la aparición de síntomas y efectos adversos, y finalmente, la disposición del paciente. Las variables de respuesta que se utilizan pueden ser agrupadas en categorías en función de la fuente de información utilizada: el propio paciente (*patient-reported outcomes*, PRO), el profesional (*clinician-reported outcomes*, CRO), pruebas de laboratorio, dispositivos, observación de la conducta y medidas indirectas (discapacidad, utilización de servicios, etc.) (Turk et al, 2006).

Cada fuente proporciona información diferente. Por ejemplo, los PRO muestran la percepción de los pacientes sobre el impacto de la enfermedad y el tratamiento sobre la salud y su capacidad funcional, y permiten medir, entre otros aspectos, su estado de salud, los síntomas, el cumplimiento del tratamiento o su satisfacción. Los CRO permiten medir resultados observados por un profesional o que requieren su interpretación (radiología, bioquímica, etc.), incluyendo escalas cumplimentadas por un profesional sanitario sobre aspectos del paciente que requieren un juicio o interpretación por parte de un clínico. Las pruebas de laboratorio y los dispositivos suelen incluir medidas objetivas cuantitativas, tanto sobre aspectos fisiológicos como conductuales, mediante protocolos de observación estructurados.

Dado que la conclusión del estudio depende del análisis de la variable de respuesta, a la hora de seleccionarla debe tenerse en cuenta que debe conducir a resultados que permitan una interpretación clínica concreta. Esta selección debe estar bien justificada y argumentada en el protocolo, y los criterios para la interpretación clínica de los resultados claramente establecidos.

**Ejemplo 13.6.** En los estudios que evalúan el efecto de una intervención sobre la presión arterial, pueden existir diferentes métodos de definición de la variable de respuesta, y debe seleccionarse como medida principal la que se considera que tiene una interpretación clínica

más relevante. Por ejemplo, puede expresarse como la disminución de las cifras de presión arterial desde el inicio del estudio, ya sea en términos absolutos o en términos relativos a las cifras basales, y esta disminución puede considerarse en términos cuantitativos o bien en categorías predefinidas, como reducción importante, moderada, leve o sin reducción. También podría definirse como la proporción de pacientes que han reducido sus cifras de presión arterial por debajo de unos niveles prefijados. Es importante que estos criterios se defian a priori, para prevenir que los resultados observados puedan condicionar la categorización e influir sobre la interpretación de los resultados.

La guía ICH sobre evaluación clínica de los nuevos fármacos antihipertensivos (ICH, 2000) establece que la variable principal de los estudios de eficacia debe ser el cambio absoluto en los valores de presión arterial al final del estudio en relación con una línea basal sin tratamiento farmacológico. El efecto sobre la presión arterial y la relación de la respuesta con las dosis debería ser caracterizado en los estudios de corta duración (4-12 semanas), en que puede utilizarse un placebo como comparación, pero deben realizarse también estudios a más largo plazo (6 meses o más) para demostrar el mantenimiento de la eficacia.

La significación estadística del cambio de la variable entre la línea basal y el final del estudio se usa para decidir si la intervención ha producido un efecto, y la de la diferencia en dicho cambio entre dos o más grupos para comparar el efecto de intervenciones alternativas. Sin embargo, es importante reconocer que significación estadística y significación clínica no son necesariamente equivalentes, de manera que un tratamiento puede producir un cambio estadísticamente significativo que puede ser interpretado como un efecto positivo, incluso tratándose de pequeñas diferencias carentes de relevancia clínica. Por ello es imprescindible establecer a priori cuál es la mínima magnitud de este cambio o diferencia que puede considerarse clínicamente relevante para poder interpretar los resultados, valor que puede denominarse mínima diferencia importante (MDI) (cuadro 13.1).

**Ejemplo 13.7.** Existen diferentes instrumentos y procedimientos desarrollados para valorar el impacto del dolor crónico sobre la calidad de vida. Uno de los más utilizados es el *Oswestry Disability Index*, diseñado para la evaluación del dolor lumbar, formado por 10 secciones, cada una de las cuales se puntuá entre 0 y 5 (puntuación máxima = 50). Se ha establecido en 15 puntos el cambio mínimo en la puntuación de los pacientes que son sometidos a una cirugía. También se ha definido una MDI de

**Cuadro 13.1 Importancia de definir la mínima diferencia importante**

- Permite concluir sobre si una intervención es efectiva o una exposición supone un riesgo
- Facilita la interpretación de los resultados a otros investigadores y clínicos
- Facilita el diseño del estudio, y en particular el cálculo del tamaño de la muestra
- Permite saber si el resultado de un paciente puede ser considerado positivo, en función de si su respuesta es superior o inferior a este valor

Modificado de Hays et al, 2005.

4 puntos al comparar las puntuaciones medias entre grupos (Manchikanti et al, 2008).

No existe una metodología estándar para establecer el valor de la MDI. Un método sencillo y ampliamente utilizado es preguntar a un grupo de expertos cuál es el valor que ellos consideran relevante.

**Ejemplo 13.8.** Le Saux et al (2005) llevaron a cabo un ensayo clínico en niños con otitis media aguda diagnosticada clínicamente, con la finalidad de demostrar que la estrategia de «ver y esperar» no era inferior a la administración de amoxicilina. La variable principal del estudio fue la resolución clínica de los síntomas, definida como la ausencia de administración de un antibiótico (diferente de la amoxicilina en el grupo con tratamiento activo) durante los 14 días posteriores, dado que la prescripción de un antibiótico se consideró un marcador de la ausencia de mejora clínica. Los autores establecieron una diferencia del 10% en la frecuencia de resolución entre ambos grupos como la mínima clínicamente relevante, de forma que una diferencia superior a este valor indicaría la inferioridad de la estrategia de «ver y esperar» frente a la otra. Para definir este límite, antes del inicio del ensayo llevaron a cabo una encuesta en médicos de urgencias del mismo centro, que coincidieron en estimar que una diferencia del 10% podría considerarse aceptable para probar la no inferioridad de la estrategia.

El principal problema de este método es que se basa exclusivamente en la opinión de un grupo concreto de expertos, que puede no coincidir con la de otros profesionales o con la de los propios participantes, por lo que es muy dependiente del contexto específico en que se ha determinado. Así, en el ejemplo anterior, probablemente se hubiera obtenido un valor diferente para la MDI si se hubiera consultado

a pediatras en lugar de a profesionales de urgencias, o a los padres de los niños afectados.

Un abordaje más adecuado sería la demostración de que un cambio en la variable de respuesta se correlaciona con un cambio relevante en el estado de salud. Por ejemplo, a partir de estudios longitudinales, podría analizarse la correlación existente entre los cambios en la presión arterial y en el riesgo cardiovascular, determinando cuál es el cambio mínimo en los valores de la presión arterial que se asocian a un cambio clínicamente relevante en el riesgo cardiovascular. Este método suele denominarse «de ancla». Otro abordaje es utilizar un método basado en la distribución estadística, relacionando los cambios en la variable con la variabilidad de sus valores. Existen diferentes métodos en función de la medida de variabilidad que se utiliza en el cálculo. Estos métodos se desarrollan con más detalle en el capítulo 21.

**Ejemplo 13.9.** Aunque se han desarrollado diferentes instrumentos para medir la respuesta de los ensayos clínicos en disnea, no existe un acuerdo sobre cómo interpretar la magnitud de las diferencias observadas. Para determinar la MDI, Ries (2005) utilizó los datos de una revisión sistemática de ensayos clínicos que evaluaban la respuesta a una intervención (rehabilitación respiratoria o ejercicio). Utilizando un método basado en la distribución estadística denominado «tamaño del efecto», estableció una MDI de cinco unidades para el *USCD Shortness of Breath Questionnaire (SOBQ)*, de una unidad para la escala Borg y de entre 10 y 20 unidades para la escala analógica visual.

**Ejemplo 13.10.** Para determinar la MDI del índice de gravedad del insomnio (Insomnia Severity Index [ISI]), Yang et al (2009) analizaron la asociación entre las diferencias en la puntuación del ISI y resultados relacionados con la salud, como la calidad de vida (medida con el cuestionario SF-36), la productividad (medida con el *Work Limitations Questionnaire, WLQ*) o la fatiga (medida con el *Fatigue Severity Scale, FSS*). Para ello, utilizaron los datos de un ensayo clínico controlado con placebo que evaluaba la eficacia a largo plazo (6 meses) de un fármaco en el insomnio primario. Aplicando modelos de regresión logística y evaluando la convergencia entre las estimaciones obtenidas por el método de ancla y por otro basado en la distribución estadística, los autores llegaron a la conclusión de que una reducción de 6 puntos en el ISI representa una mejoría clínicamente relevante en sujetos con insomnio primario, aunque recomendaban evaluar su generalizabilidad a otras poblaciones.

**Ejemplo 13.11.** Dada la incertidumbre existente sobre la interpretación de los cambios en la prueba de la distancia recorrida en 6 minutos de marcha en pacientes con enfermedad pulmonar obstructiva crónica (EPOC), Puhan et al (2008) utilizaron los datos de nueve ensayos clínicos que habían incluido un amplio abanico de pacientes con EPOC y que disponían de los resultados de la prueba para intentar establecer cuál era la MDI. Los métodos basados en la distribución estadística estimaron unos valores umbrales de relevancia clínica entre 29 y 42 m, y en concreto el basado en el error estándar de la medida condujo a un valor de 35 m (intervalo de confianza del 95% [IC 95%]: 30-42 m). Los autores consideran que este último valor podría ser la MDI en pacientes con EPOC moderada o grave, pero reconocen que su escasa correlación con las estimaciones obtenidas con el método de ancla basado en la percepción de los pacientes cuestiona la existencia de una MDI útil.

De los ejemplos anteriores se deduce que la definición de la MDI es un proceso complejo, y que los métodos disponibles para establecerla no son perfectos, aunque sí útiles, y que conducen a respuestas similares, pero no siempre coincidentes. Una buena solución es estimar la MDI utilizando diferentes procedimientos, y posteriormente realizar análisis de sensibilidad para evaluar la consistencia de los resultados. Además, la determinación de la MDI es muy dependiente del contexto concreto en que se ha determinado, por lo que hay que tener precaución al utilizarlos en entornos y poblaciones diferentes a aquellos cuyos datos se han usado para determinarla.

Por tanto, una variable de respuesta no sólo debe estar bien definida en un marco conceptual adecuado y tener una interpretación clínica concreta, sino que también debe ser capaz de detectar cambios después de una intervención o exposición o diferencias entre grupos clínicamente relevantes, es decir, debe ser capaz de detectar la MDI. Esta capacidad es conocida como sensibilidad al cambio y se comenta con más detalle en el capítulo 21.

## VARIABLES SUBROGADAS

Las *variables subrogadas* son medidas que se utilizan como sustitutos de un resultado clínicamente relevante para obtener conclusiones sobre este resultado sin tener que observarlo directamente.

Las características de una variable subrogada ideal se resumen en el cuadro 13.2. La asunción básica de su utilización es que tiene una relación causal con el resultado de interés, de manera que puede utilizarse

Cuadro 13.2 **Características de una variable subrogada ideal**

- Fiable, reproducible, clínicamente disponible, fácilmente cuantificable y mostrar un efecto «dosis-respuesta»
- Verdadera predictora de la enfermedad (o del riesgo de enfermar). Su relación con la enfermedad debe ser biológicamente plausible
- Sensible: un resultado «positivo» debe recoger a todos o la mayoría de los sujetos con incremento de riesgo de un resultado adverso
- Específica: un resultado «negativo» debe excluir a todos o la mayoría de los sujetos sin incremento de riesgo de un resultado adverso
- Punto de corte preciso entre los valores normales y anormales
- Valor predictivo positivo aceptable: un resultado «positivo» debe significar siempre o casi siempre un aumento del riesgo de un resultado adverso
- Valor predictivo negativo aceptable: un resultado «negativo» debe significar siempre o casi siempre que no existe un aumento del riesgo de un resultado adverso
- Susceptible de control de calidad
- Cambios en la variable intermedia deben reflejar rápidamente y con precisión la respuesta al tratamiento, así como normalizar sus valores en las fases de remisión o cura

Adaptada de Greenhalgh, 1997.

como predictor de la misma. La demostración de la existencia de esta relación requiere una cuidadosa evaluación de toda la evidencia disponible, tanto epidemiológica como terapéutica, patofisiológica o de cualquier tipo, lo que suele implicar la valoración de los resultados de múltiples estudios realizados con diferentes diseños. La existencia de una correlación estadística entre ambas variables es necesaria pero no suficiente. Lógicamente, cuanto más fuerte sea esta asociación, mayor es la probabilidad de que sea causal, pero la realidad es que muchas variables que se utilizan como sustitutos de los resultados de interés suelen presentar asociaciones débiles con éstos, como ocurre por ejemplo entre las medidas de función respiratoria en los pacientes con EPOC y la capacidad para realizar las actividades de la vida diaria. Además, esta asociación debería ser consistente entre diferentes estudios y haberse establecido después de ajustar por los factores de confusión conocidos, de manera que se tenga un alto grado de convicción de que la asociación existe y es independiente de otros factores.

El uso de variables subrogadas es habitual en las fases tempranas del desarrollo clínico de un fármaco, ya que se trata de determinar si puede ser potencialmente beneficioso. En fases más avanzadas, cuando la finalidad de los ensayos es confirmar la eficacia y efectividad clínicas, también podría ser beneficioso el uso de este tipo de variables para medir la respuesta terapéutica, ya que los estudios podrían ser más eficientes y contribuir a una más rápida diseminación de nuevos tratamientos eficaces, ya que los resultados se observan en un período de tiempo más corto, suelen requerir un número menor de sujetos y las variables pueden medirse con mayor precisión. Sin embargo, dado que su uso no siempre es adecuado y que los efectos sobre estas variables no siempre se corresponden con efectos similares sobre los resultados clínicos, la adopción de nuevos tratamientos basada exclusivamente en los resultados de ensayos clínicos que utilizan variables subrogadas (que, por otra parte, suponen una parte importante de los estudios que se publican en las revistas médicas) puede ser muy perjudicial por conducir a la generalización del uso de nuevos fármacos en miles de pacientes sin la suficiente evidencia de que tengan un verdadero beneficio clínico. Así, por ejemplo, hay casos de fármacos antihipertensivos y antilipemiantes que reducen la presión arterial y la colesterolemia, respectivamente, pero no el riesgo de eventos cardiovasculares, o incluso de antidiabéticos orales que reducen la hemoglobina glucosilada pero aumentan el riesgo cardiovascular.

**Ejemplo 13.12.** La rosiglitazona es ampliamente utilizada en la diabetes mellitus tipo 2. Con la finalidad de evaluar sus efectos sobre la morbilidad cardiovascular, Nissen et al (2007) realizaron una revisión sistemática de 47 ensayos clínicos que estudiaban la eficacia de dicho fármaco, y hallaron que se asociaba a un incremento significativo del riesgo de IAM (*odds ratio* [OR]=1,43; IC 95%: 1,03-1,98;  $p=0,03$ ) y a un aumento de la mortalidad de causa cardiovascular cerca del límite de la significación estadística (OR=1,64; IC 95%: 0,98-2,74;  $p=0,06$ ). Estos resultados aconsejan que se establezca el requisito de que los nuevos antidiabéticos deban demostrar que no tienen efectos perjudiciales sobre el riesgo cardiovascular.

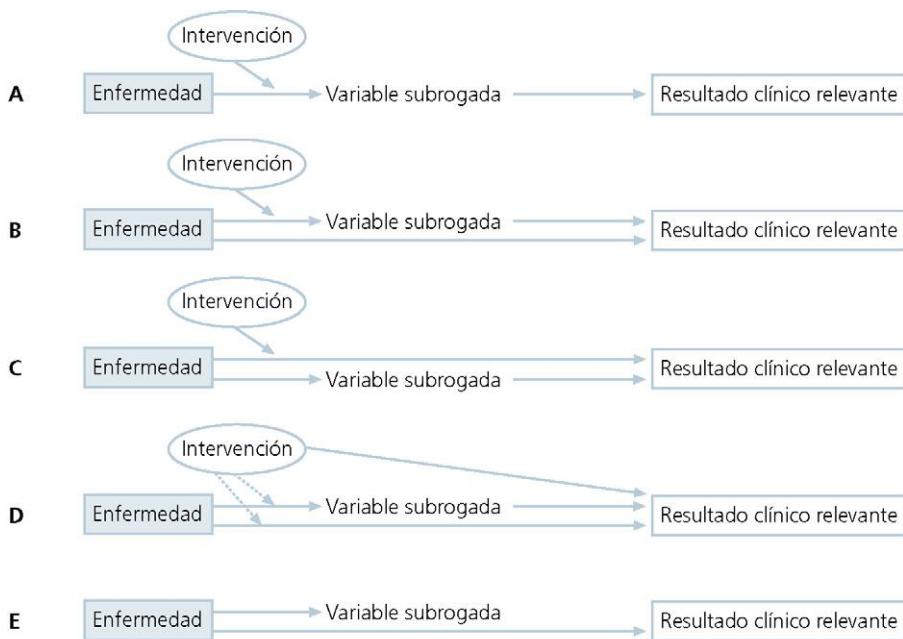
resultados de una terapia intensiva en relación con la estándar. Aunque la terapia intensiva reducía en un 21% la incidencia de microalbuminuria, tan sólo se reducía la de enfermedad renal terminal en un 5%. Por otro lado, observaron un aumento en la mortalidad total y cardiovascular en el grupo de intervención que obligó a interrumpir el estudio anticipadamente.

En el contexto de la evaluación del efecto de un tratamiento, también debe demostrarse que la variable subrogada refleja la totalidad, o como mínimo una gran parte, del efecto clínico que se evalúa. Existen múltiples situaciones en las que los efectos sobre una variable subrogada pueden no reflejarlo adecuadamente (fig. 13.1).

La situación ideal es la A, en la que la variable subrogada forma parte de la única cadena causal que conduce del efecto del tratamiento al verdadero resultado de interés, de manera que los cambios observados en ella reflejan los que se producen en la variable primaria. Sin embargo, en realidad, muchos de los resultados clínicos de interés tienen una etiología multifactorial y pueden existir diversos factores de confusión que interfieran en su relación con otras variables, de manera que existen múltiples situaciones en las que los efectos sobre una variable subrogada pueden no reflejar fielmente los efectos sobre el resultado clínico de interés.

Por ejemplo, puede ocurrir que la variable subrogada solamente refleje la actividad del tratamiento a través de un mecanismo farmacológico, pero que no indique todo el rango de acciones y efectos del medicamento, y que por tanto solamente refleje de forma parcial el verdadero efecto (situación B). Incluso podría ocurrir que existiera un efecto positivo del tratamiento tanto sobre la variable subrogada como sobre el resultado clínico, pero sin relación directa entre ellos (situación C), de manera que la variable subrogada no reflejaría adecuadamente lo que ocurre con el resultado clínico. También podría existir un efecto positivo sobre la variable subrogada, pero sin beneficio clínico (situación D), o incluso con un efecto negativo sobre el resultado clínico. Así, por ejemplo, se ha comprobado que algunos antidiabéticos orales, a pesar de que producen una disminución de la glucemia, conducen a un aumento de la morbilidad en los pacientes diabéticos. También podría ocurrir que existiera beneficio clínico sin un impacto detectable sobre la variable subrogada, ya que ésta no está realmente en la cadena causal que conduce de la intervención al resultado final (situación E). En estos casos, la medida de los cambios observados en la variable subrogada no permite cuantificar adecuadamente la magnitud del efecto sobre el resultado clínico.

**Ejemplo 13.13.** La microalbuminuria se utiliza como marcador del efecto protector del riñón de los fármacos utilizados en la diabetes y la hipertensión arterial, asumiendo que la reducción del riesgo de microalbuminuria implica la del riesgo de enfermedad renal terminal. El estudio ACCORD (Ismail-Beigi et al, 2010), realizado sobre 10.251 pacientes diabéticos, se diseñó como un ensayo clínico para evaluar los



- Situación ideal en que la variable subrogada es válida.
- Existen diferentes mecanismos causales de la enfermedad. La intervención actúa sobre el mediado por la variable subrogada.
- Existen diferentes mecanismos causales de la enfermedad. La intervención no actúa sobre el mediado por la variable subrogada.
- La intervención tiene mecanismos de acción independientes del proceso de la enfermedad.
- La variable subrogada no está en el mecanismo causal de la enfermedad.

**Figura 13.1** Posibles situaciones en la evaluación de una intervención sobre una variable subrogada y su relación con el resultado clínico de interés (*modificada de Fleming y DeMets, 1996*).

Cuando se trata de evaluar un fármaco en concreto, es importante considerar además el efecto de clase. Así, se considera más adecuado el uso de variables subrogadas para evaluar la eficacia de un nuevo fármaco si existen ensayos clínicos realizados con otros medicamentos de la misma clase que han verificado previamente la relación entre los efectos sobre la variable subrogada y el resultado clínico. Por ejemplo, dado que las tiacidas han demostrado sus efectos beneficiosos tanto sobre las cifras de presión arterial como sobre la incidencia de accidentes cerebrovasculares (ACV), la asunción de que un nuevo fármaco produce un beneficio clínico simplemente a partir de la demostración de que reduce las cifras de presión arterial es más convincente cuando se trata de una nueva tiacida que si es un medicamento de otro grupo farmacológico. Hay que reconocer, sin embargo, que la variabilidad en el concepto de clase, y las indefiniciones en sus límites, puede limitar la utilidad de este criterio. De lo que se trata en realidad es de poder asumir que la variable subrogada refleja

toda la relación entre el tratamiento y el resultado clínico, lo que es más fácil cuando el mecanismo de acción es el mismo que el de otros fármacos que lo han demostrado previamente.

**Ejemplo 13.14.** La cerivastatina es un medicamento que pertenece al grupo terapéutico de las estatinas y se utilizaba para disminuir los niveles de colesterol en sangre y prevenir la aparición de enfermedades cardiovasculares como el IAM. Fue comercializada a finales de la década de 1990 y se retiró del mercado en el año 2001 debido a la aparición de efectos secundarios graves: se produjeron 52 muertes por el uso del fármaco, debidas a rabdomiolisis que desencadenó insuficiencia renal (Furberg y Pitt, 2001).

Cuando fue comercializada en España, ya había cinco estatinas disponibles en el mercado: simvastatina, pravastatina, lovastatina, fluvastatina y atorvastatina. No todas ellas disponían de las mismas pruebas de eficacia. Solamente simvastatina y

pravastatina habían demostrado una reducción de la tasa de reinfarto y de la mortalidad en pacientes de ambos性es menores de 70 años y con cifras elevadas de colesterol. El éxito comercial de la cerivastatina se basó en que reducía más las cifras de colesterol que otras estatinas, aunque no se sabía hasta qué punto un mayor descenso relativo de la colesterolemia mejoraba más el pronóstico, ni hasta qué cifras había que reducirlo.

Parece que la dosis recomendada de cerivastatina era proporcionalmente más elevada que la de otras estatinas, lo que implicaba una mayor reducción de las cifras de colesterol pero también un aumento del riesgo de efectos adversos. De hecho, la frecuencia de aparición de rabdomiolisis, tanto letal como no letal, fue entre 18 y 80 veces más alta que con las otras estatinas.

## VARIABLES COMBINADAS

En algunas ocasiones puede ser difícil seleccionar una única variable principal de entre las posibles medidas de resultado, ya que los investigadores pueden tener interés en varias de ellas sin poder seleccionar una como la más relevante. En estas situaciones, una posible solución es el uso de una variable compuesta o combinada (*composite outcome*) como medida del resultado, englobando en un único estimador el número de pacientes que presentan al menos uno de los eventos de los componentes que incluye dicha variable. Por ejemplo, un estudio sobre prevención cardiovascular podría utilizar como variable principal de respuesta el número de episodios cardiovasculares, definido como la suma de episodios de IAM mortal, IAM no mortal, muerte súbita, ACV mortal y ACV no mortal. El uso de este tipo de variables es cada vez más frecuente en investigación clínica, ya que permite abordar el problema de la multiplicidad sin requerir ajustes por comparaciones múltiples, y tiene la ventaja adicional de que, al aumentar el número de eventos, disminuye el tamaño de la muestra necesario. Sus potenciales ventajas hacen que se abuse de su uso (cuadro 13.3).

Sin embargo, la interpretación de los resultados basados en este tipo de variables tiene importantes limitaciones, relacionadas fundamentalmente con diferencias en la importancia relativa de sus componentes (con el riesgo de «mezclar peras con manzanas»), en su frecuencia o en el efecto del tratamiento sobre cada uno de ellos, haciendo que su interpretación sea más compleja cuanto mayor es el número de elementos que la componen. Por ello, además de presentar los resultados de la variable combinada, debe informarse también de los resultados de todos

Cuadro 13.3 **Ventajas y desventajas de la utilización de variables combinadas**

### Ventajas

- Mayor número de eventos, lo que permite utilizar muestras más pequeñas o hacer un seguimiento más corto
- Evaluación de los efectos del tratamiento en diferentes procesos pero relacionados con una misma enfermedad
- Permite profundizar en los mecanismos de acción de los fármacos
- Identificación de poblaciones de pacientes que pueden beneficiarse preferentemente de la reducción de las tasas de una serie de eventos concretos
- Ayuda a extrapoluar los efectos del fármaco a múltiples enfermedades que comparten una misma patogenia

### Desventajas

- Riesgo de combinar múltiples eventos que individualmente pueden no ser significativos, o incluso con resultados que pueden ir en direcciones opuestas
- Aspectos problemáticos de definición y/o medición de cada una de las variables dentro de un mismo ensayo o entre diferentes ensayos clínicos
- Se concede el mismo peso estadístico de los diferentes componentes de la variable combinada
- Riesgo de doble recuento de alguno de los eventos si se miden todos los eventos en vez del primer evento al paciente individual
- Riesgo de que los clínicos malinterpretan los resultados al adjudicar los beneficios de la intervención a todos los componentes que integran la variable combinada, lo que sobreestimaría la percepción del efecto

los componentes de forma separada para evaluar la consistencia de los efectos o si algún componente domina sobre el resto. También debería evitarse sugerir que se ha demostrado un efecto sobre los componentes individuales, ya que el estudio se ha diseñado para detectar un posible efecto sobre dicha variable en conjunto, no sobre sus componentes. Ferreira-González et al (2008) discuten ampliamente los problemas del uso de variables combinadas.

**Ejemplo 13.15.** En el estudio UKPDS (1998) se distribuyeron 3.867 pacientes con diabetes tipo 2 en dos grupos, uno que recibió tratamiento intensivo

para el control de la glucemia y el otro tratamiento convencional. Se utilizaron tres variables de respuesta combinadas: cualquier evento relacionado con la diabetes (muerte súbita, muerte por hiperglucemias o hipoglucemias, IAM letal o no letal, angina de pecho, insuficiencia cardíaca, ACV, insuficiencia renal, amputación de al menos un dedo, hemorragia vítreas, fotoocoagulación retiniana, ceguera o extracción de cataratas), muerte relacionada con la diabetes (por IAM, ACV, enfermedad vascular periférica, enfermedad renal, hiperglucemias o hipoglucemias, y muerte súbita) y muerte por cualquier causa. En total, se estudiaron 21 componentes. También se evaluaron resultados clínicos y variables subrogadas subclínicas.

Aunque es cierto que, en relación con el grupo convencional, se observó una reducción significativa de la primera variable de resultado combinada (reducción del riesgo relativo [RRR] = 12%; IC 95%: 1-21%;  $p = 0,03$ ), la mayor parte del efecto se debió a una reducción del 25% (IC 95%: 7-40%;  $p = 0,01$ ). No se observaron reducciones estadísticamente significativas de las muertes relacionadas con la diabetes ni de la mortalidad total.

Cuando los resultados de los componentes de una variable combinada no se analizan separadamente, no puede discernirse cuál o cuáles de ellos han respondido al tratamiento, y puede interpretarse erróneamente como que el efecto se ha producido sobre todos ellos por igual.

En el cuadro 13.4 se presentan los principales elementos que se deben tener en cuenta para valorar la adecuación del uso de variables combinadas en un estudio. Idealmente, con la finalidad de evitar que alguno de los componentes tenga un mayor

#### Cuadro 13.4 Preguntas para evaluar la adecuación del uso de variables combinadas

- ¿Las diferentes variables de resultado que componen la variable principal combinada son de una importancia similar para los pacientes?
- ¿Los diferentes componentes de la variable principal tienen un sustrato biológico común?
- ¿Las variables de resultado que componen la variable principal combinada, tanto las más como las menos relevantes, ocurren con la misma frecuencia?
- ¿Los estimadores de la reducción del riesgo relativo de cada una de las variables de resultado que componen la variable principal combinada son similares y sus intervalos de confianza son lo suficientemente estrechos?

peso sobre el resultado final que el resto, todos ellos deben tener una importancia similar para el paciente, deben ocurrir con una frecuencia parecida, y presentar una magnitud del efecto también similar. Si alguna de estas condiciones no se cumple, no es adecuado combinar los distintos componentes en una única variable.

**Ejemplo 13.16.** El estudio DREAM (2006) tenía por objetivo evaluar la eficacia de la rosiglitazona en la prevención de la diabetes tipo 2 en individuos de alto riesgo de desarrollarla. Se llevó a cabo un ensayo clínico aleatorio con 5.269 sujetos donde se comparaba el fármaco con placebo. Se utilizó una variable de respuesta que combinaba la incidencia de diabetes y la mortalidad, dos componentes de diferente relevancia clínica, distinta frecuencia (alrededor del 1 y el 25%, respectivamente) y desigual efecto del tratamiento, ya que se observó una importante reducción del riesgo de diabetes, pero ningún efecto significativo sobre la mortalidad (tabla 13.2). Por tanto, en este caso, la selección de la variable combinada no fue adecuada y la interpretación de los resultados de dicha variable, sin tener en cuenta sus componentes por separado, puede llevar a confusión, ya que podría deducirse erróneamente que se reduce el riesgo tanto de la variable combinada como de todos sus componentes, cuando de hecho no refleja adecuadamente los resultados sobre ellos.

**Ejemplo 13.17.** El estudio HOPE (2000) comparó la eficacia del ramipril frente a placebo en 9.297 pacientes de alto riesgo cardiovascular. Como variable de resultado, utilizó la combinación de muerte de causa cardiovascular, IAM y ACV. A diferencia del ejemplo anterior, los tres componentes son de gran importancia clínica para los pacientes, las frecuencias con que aparecen son relativamente similares (oscilan entre el 5 y el 12% en el grupo control) y el efecto del tratamiento fue relativamente homogéneo (estimaciones del riesgo relativo entre 0,68 y 0,80) (tabla 13.3). Por tanto, la selección de la variable combinada en este caso fue adecuada.

**Ejemplo 13.18.** En un ensayo clínico en que se comparaba el efecto sobre la progresión de la nefropatía diabética de dos tratamientos activos entre sí y frente a un grupo placebo, se utilizó como medida del resultado una variable que combinaba mortalidad por todas las causas, enfermedad renal terminal y una concentración de creatinina sérica que doblara el nivel inicial, componentes de diferente relevancia clínica (Lewis et al, 2001). La tabla 13.4 presenta los

**Tabla 13.2 Resultados del estudio DREAM (2006) sobre la variable combinada y sus componentes (ejemplo 13.16)**

	<b>Intervención (n=2.365)</b>	<b>Control (n=2.634)</b>	<b>RR (IC 95%)</b>
<b>Componentes:</b>			
– Incidencia de diabetes	280 (10,6%)	658 (25,0%)	0,38 (0,33 a 0,44)
– Mortalidad	30 (1,1%)	33 (1,3%)	0,90 (0,55 a 1,50)
<b>Variable combinada</b>	306 (11,6%)	686 (26,0%)	0,40 (0,35 a 0,46)
RR: riesgo relativo. IC 95%: intervalo de confianza del 95%.			

**Tabla 13.3 Resultados del estudio HOPE (2000) sobre la variable combinada y sus componentes (ejemplo 13.17)**

	<b>Intervención (n=4.645)</b>	<b>Control (n=4.652)</b>	<b>RR (IC 95%)</b>
<b>Componentes:</b>			
– Muerte de causa cardiovascular	282 (6,1%)	377 (8,1%)	0,74 (0,64 a 0,87)
– Infarto agudo de miocardio	459 (9,9%)	570 (12,3%)	0,80 (0,70 a 0,90)
– Accidente cerebrovascular	156 (3,4%)	226 (4,9%)	0,68 (0,56 a 0,82)
<b>Variable combinada</b>	651 (14,0%)	826 (17,8%)	0,78 (0,70 a 0,86)
RR: riesgo relativo. IC 95%: intervalo de confianza del 95%.			

**Tabla 13.4 Resultados de un ensayo clínico que compara el efecto de dos tratamientos activos entre sí y frente a placebo sobre la progresión de la nefropatía diabética (Lewis et al, 2001). Se muestran los resultados de los grupos con tratamiento activo (ejemplo 13.18)**

	<b>Irbesartán (n=579)</b>	<b>Amlodipino (n=567)</b>	<b>RR (IC 95%)</b>
<b>Componentes:</b>			
– Doblar la concentración de creatinina sérica	98 (16,9%)	144 (25,4%)	0,61 (0,48 a 0,79)
– Enfermedad renal terminal	82 (14,2%)	104 (18,3%)	0,76 (0,57 a 1,02)
– Mortalidad total	87 (15,0%)	83 (14,6%)	1,05 (0,78 a 1,42)
<b>Variable combinada</b>	189 (32,6%)	233 (41,1%)	0,76 (0,63 a 0,92)
RR: riesgo relativo ajustado. IC 95%: intervalo de confianza del 95%.			

resultados de la comparación de los dos grupos que recibieron tratamiento activo, y muestran también una heterogeneidad importante en la reducción del

riesgo, ya que el irbesartán parece reducir el riesgo de los dos últimos componentes en relación con el amlodipino, pero no modifica la mortalidad total.

Así pues, la interpretación de los ensayos clínicos aleatorios que utilizan variables combinadas debe ser muy cautelosa, y la construcción de este tipo de variables requiere su definición y la de sus componentes antes del inicio del estudio, para evitar una heterogeneidad importante entre ellos tanto en términos de importancia clínica como de efecto de la intervención, por lo que se deberán ponderar si fuera necesario, y evitar también la inclusión de componentes en los que es improbable un efecto de la intervención o que carecen de importancia clínica.

## CUESTIONARIOS

Cada vez es más frecuente el uso de escalas y cuestionarios en los estudios clínicos y epidemiológicos. Como se ha comentado, cuando se utilizan como variable de respuesta, además de ser elementos válidos y fiables, han de ser sensibles a los cambios y, por tanto, capaces de detectar cambios de importancia clínica producidos en los participantes a lo largo del tiempo o diferencias relevantes entre los grupos que se comparan. No existe una única definición ampliamente aceptada de sensibilidad al cambio: Terwee et al (2003) identificaron 25 definiciones y 31 índices estadísticos para cuantificarla. Entre las definiciones más aceptadas se encuentra la de Guyatt et al (1989), que la definen como la capacidad de una escala para detectar cambios reales en el estado de salud de los participantes a lo largo del tiempo. Otras definiciones enfatizan la capacidad de la escala para detectar la diferencia mínima que se considera importante o de

relevancia clínica. Los métodos más utilizados para cuantificarla se comentan en el capítulo 21.

**Ejemplo 13.19.** Para evaluar la sensibilidad al cambio del cuestionario Fresno, diseñado para medir los conocimientos y habilidades en medicina basada en la evidencia de los profesionales sanitarios, Argimon-Pallàs et al (2011) compararon las puntuaciones del cuestionario obtenidas por los participantes antes y después de recibir una intervención educativa. Encuentran un tamaño del efecto de 1,77 (IC 95%: 1,57-1,95), y una media estandarizada del cambio de 1,65 (IC 95%: 1,47-1,82). Ambos índices indican una elevada sensibilidad al cambio del cuestionario, y las diferencias observadas, aunque no son importantes, pueden explicarse por que tienen finalidades distintas. Mientras que ambos se basan en la diferencia entre las puntuaciones antes y después de las intervenciones, existen diferencias en cómo incorporan la variabilidad en los cálculos. El tamaño del efecto usa la desviación estándar de los valores basales como denominador, y su objetivo es cuantificar la magnitud del cambio, mientras que la media estandarizada del cambio utiliza la desviación estándar de los cambios observados en las puntuaciones, por lo que su finalidad es mostrar diferencias estadísticamente significativas.

El conocimiento de la sensibilidad al cambio de una escala es muy importante para seleccionar el cuestionario más adecuado como variable de respuesta en un estudio. Además, también es útil para decidir otros aspectos del diseño de la investigación, especialmente el cálculo del tamaño de la muestra necesario.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- Argimon-Pallàs JM, Flores-Mateo G, Jiménez-Villa J, Pujol-Ribera E. Effectiveness of a short-course in improving knowledge and skills on evidence-based practice. *BMC Fam Pract.* 2011;12:64.
- CHMP (Committee for Medicinal Products for Human Use). Guideline on the evaluation of new medicinal products in the treatment of primary osteoporosis. Doc. Ref. CPMP/EWP/552/95 Rev. 2. London; 2005.
- Cooney RM, Warren BF, Altman DG, Abreu MT, Travis SPL. Outcome measurement in clinical trials for ulcerative colitis: towards standardization. *Trials.* 2007;8:17.
- D'Haens G, Sandborn WJ, Feagan BG, Geboes K, Hanauer SB, Irvine EJ, et al. Review of activity indices and efficacy points for clinical trials of medical therapy in adults with ulcerative colitis. *Gastroenterology.* 2007;132:763-86.
- DREAM. The DREAM (Diabetes REduction Assessment with ramipril and rosiglitazone Medication) Trial Investigators. Effect of rosiglitazone on the frequency of diabetes in patients with impaired glucose tolerance or impaired fasting glucose: a randomised controlled trial. *Lancet.* 2006;368:1096-105.
- Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain trials: IMMPACT recommendations. *Pain.* 2005;113:9-19.
- Furberg CD, Pitt B. Withdrawal of cerivastatin from the world market. *Curr Control Trials Cardiovasc Med.* 2001;2:205-7.
- HOPE. The Heart Outcomes Prevention Evaluation Study Investigators. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in

- high-risk patients. *N Engl J Med.* 2000;342:145-53.
- ICH (International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use). Principles for clinical evaluation of new antihypertensive drugs. Doc. E12, 2000. Disponible en: <http://www.ich.org/products/guidelines/efficacy/article/efficacy-guidelines.html>.
- Ismail-Beigi F, Craven T, Banerji MA, Basile J, Calles J, Cohen RM, et al. Effect of intensive treatment of hyperglycaemia on microvascular outcomes in type 2 diabetes: an analysis of the ACCORD randomised trial. *Lancet.* 2010;376:419-30.
- Le Saux N, Gaboury I, Baird M, Klassen TP, MacCormick J, Blanchard C, et al. A randomized, double-blind, placebo-controlled noninferiority trial of amoxicillin for clinically diagnosed acute otitis media in children 6 months to 5 years of age. *CMAJ.* 2005;172:335-41.
- Lewis EJ, Hunsicker LG, Clarke WR, Berl T, Pohl MA, Lewis JB, et al. Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. *N Engl J Med.* 2001;345:851-60.
- Manchikanti L, Hirsch JA, Smith HS. Evidence-Based Medicine, Systematic Reviews, and Guidelines in Interventional Pain Management: Part 2: Randomized Controlled Trials. *Pain Physician.* 2008;11:717-73.
- Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med.* 2007;356:2457-71.
- Puhan MA, Mador MJ, Held U, Goldstein R, Guyatt GH, Schünemann HJ. Interpretation of treatment changes in 6-minute walk distance in patients with COPD. *Eur Respir J.* 2008;32:637-43.
- Ries AL. Minimally clinically important difference for the UCSD Shortness of Breath Questionnaire, Borg Scale, and Visual Analog Scale. *COPD.* 2005;2:105-10.
- Shaughnessy AF, Slawson DC. What happened to the valid POEMs? A survey of review articles on the treatment of type 2 diabetes. *BMJ.* 2003;327:266-9.
- Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzera L. OMERACT: an international initiative to improve outcome measurement in rheumatology. *Trials.* 2007;8:38.
- Turk DC, Dworkin RH, Burke LB, Gershon R, Rothman M, Scott J, et al. Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. *Pain.* 2006;125:208-15.
- UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet.* 1998;352:837-53.
- Yang M, Morin CM, Schaefer K, Wallenstein GV. Interpreting score differences in the Insomnia Severity Index: using health-related outcomes to define the minimally important difference. *Curr Med Res Opin.* 2009;25:2487-94.

## BIBLIOGRAFÍA

- Baker SG, Kramer BS. A perfect correlate does not a surrogate make. *BMC Med Res Methodol.* 2003;3:16.
- Bethel MA, Holman R, Haffner SM, Califff RM, Huntsman-Labed A, Hua TA, et al. Determining the most appropriate components for a composite clinical trial outcome. *Am Heart J.* 2008;156:633-40.
- Brotons C. Uso de variables combinadas en ensayos clínicos. *FMC.* 2006;13:485-6.
- Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA, for the Evidence-Based Medicine Working Group. Users' Guides to the Medical Literature XIX. Applying Clinical Trial Results. *A. How to use an Article Measuring the Effect of an Intervention on Surrogate End Points. JAMA.* 1999;282:771-8.
- Clark M. Standardising outcomes for clinical trials and systematic reviews. *Trials.* 2007;8:39.
- Cooper R, Kaanders JH. Biological surrogate end-points in cancer trials: potential uses, benefits and pitfalls. *Eur J Cancer.* 2005;41:1261-6.
- Cordoba G, Schwartz L, Woloshin S, Bae H, Götzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ.* 2010;341:c3920.
- Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* 2003;56:395-407.
- De Gruttola VG, Clax P, DeMets DL, Downing GJ, Ellenberg SS, Friedman L, et al. Considerations in the evaluation of surrogate endpoints in clinical trials. Summary of a National Institutes of Health workshop. *Control Clin Trials.* 2001;22:485-502.
- Duffy SW, Treasure FP. Potential surrogate endpoints in cancer research – some considerations and examples. *Pharm Stat.* 2011;10:34-9.
- © Elsevier. Fotocopiar sin autorización es un delito.

- Ferreira-González I, Alonso-Coello P, Solà I, Pacheco-Huergo V, Domingo-Salvany A, Alonso J, et al. Variables de resultado combinadas en los ensayos clínicos. *Rev Esp Cardiol.* 2008;61:283-90.
- Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being mislead? *Ann Intern Med.* 1996;125:605-13.
- Freemantle N, Calvert M. Composite and surrogate outcomes in randomised controlled trials. *BMJ.* 2007;334:756-7.
- Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA.* 2003;289:2554-9.
- Greenhalgh T. How to read a paper: papers that report drug trials. *BMJ.* 1997;315:480-3.
- Grimes DA, Schulz KF. Surrogate end points in clinical research: hazardous to your health. *Obstet Gynecol.* 2005;105(5 Pt 1):1114-8.
- Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol.* 1989;42:403-8.
- Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD.* 2005;2:63-7.
- Huque MF, Alesh M, Bhore R. Addressing multiplicity issues of a composite endpoint and its components in clinical trials. *J Biopharm Stat.* 2011;21:610-34.
- Jiménez Villa J. Variables intermedias o subrogadas en los ensayos clínicos. *FMC.* 2006;13:331-2.
- Kaplan RM. The minimally clinically important difference in generic utility-based measures. *COPD.* 2005;2:91-7.
- Make B. How can we assess outcomes of clinical trials: the MCID approach. *COPD.* 2007;4:191-4.
- Montori VM, Permanyer-Miralda G, Ferreira-González I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. *BMJ.* 2005;330:594-6.
- Norman GR. The relation between the minimally important difference and patient benefit. *COPD.* 2005;2:69-73.
- Norman GR, Sridhar FG, Guyatt GH, Walter SD. Relation of distribution- and anchor-based approaches in interpretation of changes in health-related quality of life. *Med Care.* 2001;39:1039-47.
- Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61:102-9.
- Sloan JA. Assessing the minimally clinically significant difference: scientific considerations, challenges and solutions. *COPD.* 2005;2:57-62.
- Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res.* 2003;12:349-62.
- Wittkop L, Smith C, Fox Z, Sabin C, Richert L, Aboulker JP, et al. Methodological issues in the use of composite endpoints in clinical trials: examples from the HIV field. *Clin Trials.* 2010;7:19-35.
- Yudkin JS, Lipska KJ, Montori VM. The idolatry of surrogate. *BMJ.* 2011;343:d7995.

## Población de estudio

En la fase de diseño de una investigación se plantean tres cuestiones relacionadas con los participantes: definir la población de estudio (especificar los criterios de selección), calcular el tamaño de la muestra necesario y decidir el método para obtener la muestra (técnica de muestreo). Este capítulo se centra en la primera cuestión. Las otras dos se abordan en los capítulos siguientes.

La formulación del objetivo lleva implícita una definición genérica de la población que se desea estudiar, pero en la redacción del protocolo es necesario establecer con mayor precisión unos criterios de selección de los sujetos que se van a incluir en el estudio.

### NIVELES DE POBLACIÓN

Pueden tomarse en consideración tres niveles de población (fig. 14.1):

1. *Población diana*: conjunto de individuos al que hace referencia la pregunta principal u objetivo del estudio. Es la población a la que se desearía generalizar los resultados. Se define principalmente por sus características clínicas y demográficas generales.
2. *Población de estudio*: subconjunto de la población diana al que se tiene la intención de estudiar. Se define por los criterios de selección establecidos en el protocolo y presenta determinadas características geográficas y temporales que la hacen accesible a los investigadores.
3. *Muestra*: conjunto de individuos realmente estudiados. En la mayoría de las ocasiones, el número de sujetos necesarios para la realización del estudio es mucho menor que el de candidatos que forman la población de estudio, por lo que, por razones de eficiencia y disponibilidad de recursos (viabilidad),

se selecciona una muestra. En otras ocasiones se incluyen sujetos consecutivamente hasta alcanzar el número necesario. En cualquier caso, la estrategia de inclusión de los sujetos debe intentar que la muestra sea representativa de la población de estudio.

**Ejemplo 14.1.** Se desea ensayar un nuevo diurético para el control de la hipertensión arterial (HTA) esencial leve-moderada en pacientes de la tercera edad. La *población diana* serían, pues, los sujetos de la tercera edad con HTA esencial leve-moderada. A continuación,



Figura 14.1 Selección de los sujetos de estudio.

deben establecerse criterios de inclusión y exclusión para definir la *población de estudio*: sujetos mayores de 65 años, con HTA esencial, con cifras de presión arterial diastólica (PAD) entre 90 y 109 mmHg, sin tratamiento previo, en los que se ha seguido un protocolo para descartar HTA secundaria, que no han recibido tratamiento antihipertensivo previamente, que no presentan otras patologías concomitantes y que han sido visitados y diagnosticados en los centros en que trabajan los investigadores. Se ha calculado que son necesarios 200 sujetos para la realización del estudio, por lo que los investigadores incluyen a los primeros 200 sujetos que cumplen los criterios de selección (*muestra*).

La consideración de estos niveles tiene repercusiones sobre la interpretación de los resultados. Un investigador desea que sus observaciones puedan generalizarse a todos los individuos que forman la población diana, independientemente de que se visiten o no en los centros en que se realiza el estudio. La posibilidad de realizar inferencias sobre otras poblaciones dependerá, en primer lugar, de cómo se haya diseñado y realizado la investigación, de si la muestra es representativa de la población de estudio, de si se han perdido sujetos que puedan comprometer esta representatividad, de si los datos han sido medidas y registrados correctamente, y de otros factores que condicionan la validez interna. Por otro lado, la validez externa, o capacidad de extrapolación de los resultados a poblaciones diferentes o su generalización a otras más amplias, depende de factores como las diferencias entre la población de estudio y la población diana o entre el tipo de sujetos que se visitan en los centros de estudio y la población diana.

**Ejemplo 14.2.** Van Spall et al (2007) revisaron sistemáticamente los ensayos clínicos publicados desde 1994 hasta 2006 en revistas con un alto factor de impacto, con el fin de conocer la naturaleza de los criterios de exclusión. Los autores observaron que el ser mujer, niño o anciano son criterios de exclusión frecuentes. También observaron que los criterios de inclusión son más restrictivos en los ensayos multicéntricos y en los que se evalúa un tratamiento farmacológico.

## PRINCIPIOS DE REPRESENTATIVIDAD Y COMPARABILIDAD

La muestra de sujetos estudiada debe ser representativa de la población de estudio (*principio de representatividad*). Este principio puede verse comprometido

cuando la muestra inicial ha sido mal seleccionada o cuando, aun habiendo utilizado una técnica de muestreo adecuada, la variabilidad aleatoria (el azar) ha hecho que se obtenga una muestra no representativa, o cuando la muestra de sujetos finalmente analizados está sesgada debido a las no respuestas (sujetos de la muestra inicial de los que no se ha podido obtener la información deseada) o pérdidas durante el seguimiento (en estudios prospectivos).

**Ejemplo 14.3.** Supóngase que se desea conocer la opinión de los médicos de una provincia sobre un determinado tema y, para ello, se envía un cuestionario a profesionales que trabajan en equipos de atención primaria (EAP). La muestra resultante no es representativa de los médicos de la provincia, ya que no incluye a los médicos que no trabajan en EAP ni a los que sólo tienen consulta privada, por ejemplo.

**Ejemplo 14.4.** Supóngase que en el ejemplo anterior se selecciona una muestra aleatoria de médicos a partir del registro del Colegio de Médicos de la provincia, a la que se le envía un cuestionario. Responde el 40% de los profesionales. Si la opinión de los que no responden difiere de la de los que sí lo han hecho, el resultado basado únicamente en las respuestas obtenidas estará sesgado, ya que la muestra analizada no es representativa de la población de estudio.

En los estudios analíticos debe cumplirse, además, el *principio de comparabilidad* de los grupos. Estos estudios se basan en que los grupos son comparables por todos los factores pronóstico y en que se ha obtenido la información de la misma forma en todos los sujetos, de manera que las diferencias en los resultados observados pueden atribuirse al factor que se está estudiando.

**Ejemplo 14.5.** En un estudio se revisan las historias clínicas de los sujetos que han sufrido una enfermedad con la finalidad de comparar su pronóstico en función de si han recibido tratamiento médico o quirúrgico. Ya que la indicación del tipo de tratamiento depende de la gravedad de la enfermedad y de otros factores, los grupos no son comparables (p. ej., entre los sujetos que han recibido tratamiento médico pueden haber tanto sujetos con enfermedad leve como individuos que estén tan graves que en ellos esté contraindicada la cirugía).

El principio de comparabilidad se discute más ampliamente en el capítulo destinado a la formación de los grupos de estudio.

## POBLACIÓN DE ESTUDIO

Su definición permitirá identificar en qué individuos se evaluará la presencia del efecto o la asociación de interés, evaluar su idoneidad para alcanzar el objetivo, así como la posibilidad de extrapolación de los resultados a otras poblaciones. Así mismo, permitirá que otros investigadores puedan utilizar criterios semejantes a la hora de diseñar estudios con objetivos similares.

### Características generales

La población de estudio debe definirse con criterios precisos, teniendo en cuenta las siguientes consideraciones (**cuadro 14.1**):

- *Seleccionar pacientes idóneos para observar el efecto o la asociación de interés.* Para identificar las características de los sujetos en que es más plausible encontrar el efecto o la asociación de interés, hay que tener en consideración el mecanismo de acción del factor de estudio. Por ejemplo, en un estudio observacional sobre la asociación entre consumo de anticonceptivos orales y cáncer, se incluirían –obviamente– mujeres en edad fértil y se excluirían aquellas a las que se ha practicado una histerectomía.
- En los estudios experimentales sobre el efecto de una intervención, serían candidatos los pacientes que puedan beneficiarse de ella, y se excluirían los que presenten enfermedades o situaciones concomitantes (embarazo, edad avanzada, alergias, etc.) en las que la intervención pueda resultar peligrosa o causar efectos indeseables.
- *Seleccionar individuos en los que exista una alta probabilidad de detectar el efecto o la asociación de interés.* La selección de los sujetos debe hacerse de forma que se optimicen las posibilidades de detectar la diferencia o asociación que se busca.

Cuadro 14.1 Características generales de la población de estudio

- Pacientes idóneos para observar el efecto o asociación de interés
- Individuos en los que exista una alta probabilidad de detectar el efecto hipotético o la asociación de interés
- Sujetos que probablemente cumplirán el protocolo de estudio (en función del objetivo)
- Excluir aquellos sujetos en los que exista una alta probabilidad de que no pueda medirse la respuesta

Esta consideración es especialmente importante cuando se estudian fenómenos que se presentan con una baja frecuencia o con una latencia muy prolongada. Para reducir el número de sujetos necesarios o la duración del seguimiento, muchos estudios seleccionan poblaciones de alto riesgo.

**Ejemplo 14.6.** El *Heart Protection Study* (2002) tuvo como objetivo evaluar si un tratamiento hipolipemiante podría reducir la morbilidad y la mortalidad cardiovascular en pacientes de alto riesgo, independientemente de sus cifras de colesterol. Se incluyeron 20.536 hombres y mujeres de 40 a 80 años de edad, con unas concentraciones plasmáticas de colesterol total de al menos 3,5 mmol/l, siempre que se considerara que tenían, en un plazo de 5 años, un riesgo considerable de muerte a causa de unos antecedentes previos de enfermedad coronaria, enfermedad oclusiva de las arterias no coronarias, diabetes, HTA tratada o una combinación de estos factores. Esta muestra fue distribuida aleatoriamente en dos grupos: uno de ellos recibió un tratamiento hipolipemiante, y el otro un placebo. La intención de los investigadores era seleccionar un grupo de muy alto riesgo, ya que, si no, se hubiera necesitado un tamaño de muestra y un seguimiento mucho mayor para encontrar un efecto significativo.

• *Seleccionar sujetos que probablemente cumplirán el protocolo de estudio.* Se pretende evitar situaciones como omisiones de información o incumplimientos de la intervención que podrían comprometer la validez interna del estudio.

**Ejemplo 14.7.** En un estudio que evaluaba la efectividad de una técnica de educación sanitaria para disminuir las cifras tensionales de los pacientes con HTA esencial de un centro de salud (Mata et al, 1987), se escogieron 113 individuos que cumplían los criterios de selección. De ellos, se excluyeron, entre otros, los siguientes: 10 pacientes que no habían acudido a las citaciones en los últimos 6 meses, ya que muy posiblemente tampoco cumplirían el protocolo del estudio; un paciente que presentaba hipoacusia importante, dado que el método de educación sanitaria consistía en charlas que difícilmente habría podido seguir; dos casos por enolismo crónico, ya que es conocido que por regla general son malos cumplidores, especialmente de medidas higienicodegéticas; un caso de analfabetismo, ya que se repartían documentos escritos de apoyo a la intervención; tres casos por incapacidad física para desplazarse al centro, y dos por demencia.

Sin embargo, la selección de sujetos muy cumplidores tiene el inconveniente de que aleja la población de estudio de la diana, limitando la generalización de resultados.

- *Excluir sujetos en los que exista una alta probabilidad de que no pueda ser medida la respuesta*, ya que podrían dificultar la interpretación de los resultados. Esta consideración es especialmente importante en los estudios con un seguimiento prolongado, durante los cuales los sujetos pueden presentar otros procesos que impidan su seguimiento o la medición adecuada de la respuesta.

**Ejemplo 14.8.** En el ejemplo 14.7, los autores excluyeron, además de los ya citados, a un paciente en situación administrativa de desplazado, ya que probablemente no completaría el estudio, y un caso de neoplasia terminal. A menudo, puede ser conveniente excluir pacientes embarazadas, ya que el parto puede interferir en el seguimiento.

### Criterios de selección

Los criterios de selección deben establecerse en términos realistas que permitan la inclusión del número deseado de sujetos en el tiempo previsto. Aunque es posible que sean diferentes en estudios con objetivos similares, pueden agruparse en cuatro grandes categorías (**cuadro 14.2**):

1. *Características sociodemográficas*. La edad, el sexo, la profesión o los años de escolaridad pueden ser fundamentales para definir las características de los participantes y las limitaciones que puede haber para la extrapolación a otras poblaciones.
2. *Características de la enfermedad o exposición*. Es el criterio de inclusión más importante. El diagnóstico de una enfermedad o la identificación de una exposición deben establecerse con la mayor precisión posible para evitar que puedan existir dudas acerca de si un sujeto cumple o no este criterio. En la medida de lo posible es conveniente utilizar criterios estándar ampliamente aceptados o semejantes a los de otros estudios, con la finalidad de facilitar la comparación de los resultados.

**Ejemplo 14.9.** Marantz et al (1988) revisaron los ensayos clínicos aleatorios de fármacos frente a placebo en la insuficiencia cardíaca congestiva, realizados a doble ciego y publicados entre 1977 y 1985, con la finalidad de evaluar la homogeneidad de los criterios diagnósticos utilizados. Sólo 23 de los 51 estudios revisados (45%) especificaban dichos criterios, y éstos hacían referencia tanto a síntomas y signos clínicos como

Cuadro 14.2 **Factores que se deben considerar en la definición de los criterios de selección de los sujetos**

#### Características sociodemográficas de los pacientes

- Edad
- Sexo
- Raza
- Nivel cultural
- Estado socioeconómico
- Profesión
- Otras

#### Características de la enfermedad o exposición

- Definición de enfermedad o exposición
- Características: forma, tipo, estadio, curso clínico, etiología, complicaciones, duración, etc.
- Tratamiento actual
- Tratamientos previos
- Otras enfermedades o exposiciones
- Otros tratamientos

#### Otras características

- Embarazo o lactancia
- Consumo de tabaco o alcohol
- Abuso de drogas
- Hábitos dietéticos o nutricionales
- Factores de riesgo
- Limitaciones fisiológicas
- Características psicológicas
- Hipersensibilidad a tratamientos o exploraciones
- Factores ambientales
- Participación en otros estudios
- Incapacidades

#### Características de accesibilidad

- Lugar de residencia
- Visitado en el centro sanitario
- Inicio y final del período de inclusión

Esta lista es orientativa, y no siempre es preciso considerar todas las características. En algunas ocasiones deben valorarse otros factores.

a datos de laboratorio en combinaciones diferentes. Sólo cuatro estudios presentaban criterios idénticos, y todos ellos correspondían a estudios realizados por el mismo equipo investigador. Por otro lado, las diferencias en los criterios utilizados podrían explicar algunos de los resultados contradictorios observados.

Es conveniente tener presente que la inclusión de falsos positivos o falsos negativos suele conducir a una dilución del efecto, es decir, disminuye la probabilidad de detectarlo. También es importante establecer criterios relacionados con sus características clínicas de tipo, evolución, gravedad, complicaciones y tratamiento. Además, hay que considerar otras enfermedades, procesos o tratamientos concomitantes que puedan ser de interés como criterios de inclusión o de exclusión.

- 3. Características de accesibilidad de la población.** No toda la población definida por las características de los apartados anteriores es accesible al investigador. Por ello, deben fijarse criterios de accesibilidad geográfica (lugar de residencia, atención en determinado centro sanitario, etc.) y temporal (inicio y final del período de inclusión).
- 4. Otras características.** En determinadas situaciones, la existencia de embarazo o lactancia, o de alcoholismo u otras drogodependencias, o incluso la participación del paciente en otros estudios pueden aconsejar su exclusión.

De esta forma, los criterios establecidos permiten seleccionar un conjunto de sujetos. Este planteamiento es útil cuando se va a estudiar una única muestra de sujetos para estimar un parámetro o medir un efecto en un solo grupo de individuos (como ocurre en los estudios transversales que estiman una prevalencia o en los longitudinales que estiman una incidencia), o cuando se va a subdividir la muestra en dos grupos en función de su exposición a un factor de riesgo (como en los estudios de cohortes con control interno) o de forma aleatoria para comparar dos intervenciones (como en los ensayos clínicos aleatorios).

Sin embargo, otros diseños, como los estudios de casos y controles o los de cohortes con control externo, requieren la definición de criterios específicos para el grupo control. Por ejemplo, en los estudios de casos y controles hay que definir qué patologías se incluyen en el grupo control y, a la vez, excluir aquellos sujetos que puedan estar relacionados positiva o negativamente con el factor de estudio.

### Criterios de selección amplios o restrictivos

A la hora de definir los criterios de selección, los investigadores pueden adoptar diferentes perspectivas. En un extremo se sitúa la utilización de criterios restrictivos, que definan una población de estudio muy homogénea y cumplidora con el protocolo, que sea fácil de estudiar y en la que existan mayores posibilidades

de detectar una diferencia o asociación, si existe. Esta postura tiene el inconveniente de que la población de estudio difiere enormemente de la diana, por lo que los resultados serán difícilmente generalizables o extrapolables. En el otro extremo se sitúa la definición de criterios amplios, que buscan facilitar la generalización de los resultados definiendo una población de estudio más representativa de la diana, aun a costa de que la muestra de sujetos sea muy heterogénea y se dificulte la detección de un efecto.

**Ejemplo 14.10.** Kennedy et al (2003) compararon los sujetos asmáticos tratados con corticoides inhalados que habían participado en ensayos clínicos con los que no lo habían hecho, y comprobaron que los primeros tomaban dosis superiores de corticoides y acudían menos a los servicios de urgencias, lo que demuestra la existencia de diferencias que pueden limitar la generalización de los resultados.

Al definir los criterios de selección, los investigadores deben tener en cuenta algunas consideraciones generales (cuadro 14.3) y sopesar la necesidad de definir una población lo suficientemente homogénea como para permitir detectar el efecto o la asociación que se busca con la necesidad de facilitar la generalización de los resultados, lo que habitualmente implica adoptar una posición intermedia entre ambos extremos.

### Procedencia de los sujetos

La procedencia de los sujetos ha de estar en consonancia con el objetivo. Si el interés está en conocer la prevalencia de una enfermedad en la comunidad, deberá escogerse una muestra representativa de la

**Cuadro 14.3 Consideraciones generales en la definición de criterios para la selección de la población de estudio**

- Optimización de las posibilidades de detectar una diferencia o asociación, si existe
- Población suficientemente homogénea respecto a las características más relevantes del fenómeno de estudio
- Debe asemejarse, dentro de unos límites razonables, a la población diana a la que se desearía generalizar los resultados
- Criterios realistas que permitan la inclusión del número deseado de sujetos en el tiempo previsto
- Aspectos éticos

población de dicha comunidad. Si se optara por una muestra de individuos atendidos en un centro sanitario, se obtendría un resultado sesgado, ya que la demanda de servicios estaría influida, además de por las necesidades biológicas, por factores como el sistema sanitario, los propios profesionales o factores socioculturales, entre otros. En cambio, si el objetivo es conocer la prevalencia de esta misma enfermedad en la población demandante de un centro, la muestra deberá ser representativa de aquellos que se visitan en el centro.

Desde un punto de vista estricto, los resultados de un estudio son aplicables a la población de la que la muestra de sujetos estudiados sea realmente representativa. Por tanto, los resultados que se obtienen pueden variar en función de los criterios de selección utilizados y del ámbito o tipo de centro en que se realiza el estudio. De hecho, algunas de las diferencias observadas en la literatura entre estudios dependen más de la diferente procedencia de los sujetos que de otros aspectos del diseño.

**Ejemplo 14.11.** Ellenberg y Nelson (1980) revisaron los estudios publicados sobre la pertinencia del tratamiento preventivo de las recurrencias de las crisis convulsivas febres en la infancia. Por un lado, en los estudios de base poblacional se encontraron tasas de recurrencia muy bajas, de entre el 1,5 y el 4,6%, mientras que si los niños estudiados procedían de consultas hospitalarias o unidades especializadas, las tasas eran muy variables, y en general más altas (desde el 2,6 hasta el 76,9%). Si se tuvieran en cuenta únicamente los resultados del último tipo de estudios, la necesidad de un tratamiento preventivo de las recurrencias parecería clara. Sin embargo, si se consideran los estudios de base poblacional, esta decisión no es tan evidente. Esto no significa que unos sean mejores que otros, sino simplemente que las poblaciones estudiadas son distintas en lo que se refiere al tipo de enfermedad, su gravedad, otras patologías concomitantes, etc., por lo que es importante que los investigadores utilicen criterios de selección coherentes con el objetivo concreto del estudio.

## INCLUSIÓN DE LOS SUJETOS

El reclutamiento con éxito de los participantes es esencial en toda investigación. La estrategia de inclusión de sujetos busca obtener una muestra representativa de la población definida por los criterios de selección, del tamaño suficiente y de forma eficiente.

**Ejemplo 14.12.** Moore et al (2000) analizaron las características de los participantes en dos grandes ensayos sobre tratamiento antirretroviral en sujetos infectados por el VIH, y observaron que existía una infrarepresentación de pacientes que se habían infectado por vía heterosexual o por el uso de drogas por vía parenteral, una mayor frecuencia de pacientes homosexuales y una menor probabilidad de pérdidas de seguimiento al año, en relación con los que cumplían los criterios de selección, pero que no fueron incluidos en el estudio.

El éxito del reclutamiento depende de diversos factores, entre los que destacan los relacionados con el propio fenómeno de estudio, las características socioculturales de la población y su relación con los profesionales o servicios sanitarios, y la forma en que se solicite el consentimiento para participar. Por ello, al solicitar la colaboración es importante presentar de forma comprensible el tema y los objetivos del estudio, la importancia de su participación, la institución que avala el proyecto, la garantía de confidencialidad de la información, los beneficios potenciales y los posibles riesgos y molestias que podría suponerles su participación, etc.

Cuando se diseña la estrategia de inclusión de los sujetos, deben preverse las posibles causas que puedan reducirla y establecer mecanismos para evitarlas. En estudios de larga duración es frecuente que la tasa de reclutamiento disminuya como consecuencia del cansancio de los investigadores. Suele ser útil monitorizar la tasa de inclusión a medida que el estudio progresá, para detectar la aparición de estos problemas y poder solucionarlos.

## ESTRATEGIAS PARA DISMINUIR LAS NO RESPUESTAS Y LAS PÉRDIDAS DE SEGUIMIENTO

La representatividad de la muestra debe garantizarse a lo largo de todo el proceso de la investigación de forma que se eviten sesgos en el análisis de los resultados.

En los diseños transversales es importante conseguir el mayor porcentaje de respuestas posible. Si el reclutamiento es telefónico, deben preverse múltiples llamadas en diferentes franjas horarias, y se deben incluir también los fines de semana. Si es por carta, por ejemplo, puede enviarse un agradecimiento a los que han respondido, un recordatorio a la semana de que haya vencido el plazo de participación a aquellos que no han respondido, una nueva carta a las tres semanas si todavía no lo han hecho,

y, si es necesario, una tercera carta, aproximadamente a las siete u ocho semanas de la primera.

En los estudios que requieren seguimiento, deben preverse estrategias para reducir el número de pérdidas. En los criterios de selección, deben excluirse los individuos que es probable que se pierdan durante el seguimiento, por ejemplo, porque piensan cambiar de domicilio. Es conveniente recoger

información que permita localizar a los individuos en caso de que interrumpan el seguimiento (p. ej., la dirección y el teléfono del individuo, los amigos y los familiares), así como mantener contactos periódicos con ellos, telefónicamente o por correo, que también servirán para determinar con mayor exactitud el momento en que aparecen los desenlaces de interés.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- Ellenberg JH, Nelson KB. Sample selection and the natural history of disease: studies of febrile seizures. *JAMA*. 1980;243:1337-40.
- Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet*. 2002;360:7-22.
- Kennedy WA, Laurier C, Malo JL, Ghezzo H, L'Archeveque J, Contandriopoulos AP. Does
- Marantz PR, Alderman MH, Tobin JN. Diagnostic heterogeneity in clinical trials for congestive heart failure. *Ann Intern Med*. 1988;109:55-61.
- Mata M, Monraba M, Travería M, Guarner MA, Antó JM. Ensayo aleatorio controlado de educación sanitaria en pacientes hipertensos en asistencia
- clinical trial subject selection restrict the ability to generalize use and cost of health services to «real life» subjects? *Int J Technol Assess Health Care*. 2003;19:8-16.
- primaria. *Aten Primaria*. 1987;4:189-94.
- Moore DA, Goodall RL, Ives NJ, Hooker M, Gazzard BG, Easterbrook PJ. How generalizable are the results of large randomized controlled trials of antiretroviral therapy? *HIV Med*. 2000;1:149-54.
- Van Spall HG, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA*. 2007;297(11):1233-40.

## BIBLIOGRAFÍA

- Akl EA, Gaddam S, Mustafa R, Wilson MC, Symons A, Grifasi A, et al. The effects of tracking responses and the day of mailing on physician survey response rate: three randomized trials. *PLoS One*. 2011;6(2):e16942.
- Blümle A, Meerpohl JJ, Rücker G, Antes G, Schumacher M, von Elm E. Reporting of eligibility criteria of randomised trials: cohort study comparing trial protocols with subsequent articles. *BMJ*. 2011;342:d1828.
- Edwards PJ, Roberts I, Clarke MJ, Diguiseppi C, Wentz R, Kwan I, et al. Methods to increase response to postal and electronic questionnaires. *Cochrane Database Syst Rev*. 2009;(3):MR000008.
- Graffy J, Bower P, Ward E, Wallace P, Delaney B, Kinmonth AL, et al. Trials within trials? Researcher, funder and ethical perspectives on the practicality and acceptability of nesting trials of recruitment methods in existing primary care trials. *BMC Med Res Methodol*. 2010;10:38.
- Page MJ, French SD, McKenzie JE, O'Connor DA, Green SE. Recruitment difficulties in a primary care cluster randomised trial: investigating factors contributing to general practitioners' recruitment of patients. *BMC Med Res Methodol*. 2011;11:35.
- Treweek S, Mitchell E, Pitkethly M, Cook J, Kjeldstrøm M, Taskila T, et al. Strategies to improve recruitment to randomised controlled trials. *Cochrane Database Syst Rev*. 2010; (1):MR000013.
- Williamson MK, Pirkis J, Pfaff JJ, Tyson O, Sim M, Kerse N, et al. Recruiting and retaining GPs and patients in intervention studies: the DEPS-GP project as a case study. *BMC Med Res Methodol*. 2007;7:42.

## Tamaño de la muestra

El cálculo del tamaño de la muestra permite responder a las siguientes preguntas: ¿cuántos individuos es necesario estudiar para poder estimar un parámetro determinado con el grado de confianza deseado? ¿Cuántos sujetos es necesario estudiar para tener las mínimas garantías de poder detectar una determinada diferencia entre los grupos de estudio, en el supuesto de que esta diferencia exista realmente?

El resultado del cálculo debe considerarse orientativo, ya que se basa en asunciones que pueden ser incorrectas. Por ello, es importante realizarlo cuidadosamente, ya que tiene implicaciones sobre la viabilidad del estudio y puede suponer modificaciones en su diseño (modificación de los criterios de selección, prolongación del período de inclusión, etc.).

La inclusión de un número excesivo de sujetos encarece el estudio tanto desde el punto de vista económico como de los recursos humanos y físicos. Además, puede ser considerado poco ético, ya que un mayor número de pacientes estará sometido innecesariamente a determinadas exploraciones o recibirá la intervención menos eficaz. Por otra parte, un estudio con un tamaño insuficiente de la muestra estimará un parámetro con poca precisión o será incapaz de detectar diferencias entre los grupos, conduciendo erróneamente a la conclusión de que no existen.

En general, puede considerarse que el objetivo de una investigación pertenece a uno de los dos tipos siguientes: *estimación de un parámetro* (p. ej., el porcentaje de pacientes hipertensos que presentan una determinada complicación) o *contraste de hipótesis* (p. ej., evaluar cuál de dos pautas terapéuticas consigue un mayor porcentaje de éxitos).

### ESTUDIOS QUE ESTIMAN UN PARÁMETRO

Se trata de estudios en los que los investigadores, partiendo de los valores observados en la muestra estudiada, y utilizando la inferencia estadística, buscan estimar el valor de un parámetro en la población.

### Consideraciones previas: intervalo de confianza

Supongamos que se desea estimar el valor medio de la colesterolemia en una población. Se escoge una muestra de 80 sujetos en los que se observa una media de 180 mg/100 ml, con una desviación estándar de 20 mg/100 ml. Si se hubiera estudiado una muestra diferente, seguramente se habrían obtenido cifras distintas aunque los criterios de selección hubiesen sido los mismos, si bien es probable que el valor observado en esta segunda muestra hubiera estado alrededor de 180 mg/100 ml. Por lo tanto, a partir del valor observado en una única muestra no puede conocerse exactamente el verdadero valor en la población de origen, ya que es tan sólo una aproximación (estimación puntual). Para conocer entre qué límites es más probable que se sitúe este verdadero valor, debe calcularse el *intervalo de confianza* (IC).

El método de cálculo del IC en las situaciones más frecuentes se desarrolla en los capítulos dedicados al análisis de los datos. La amplitud del IC, es decir, la precisión de la estimación, depende del grado de confianza, la variabilidad del parámetro de interés y el número de sujetos estudiados. Cuanto menor sea la variabilidad del parámetro y mayor el número

de sujetos, mayor precisión existirá en la estimación para un grado de confianza determinado. Cuanto más confianza se desee tener, más amplio será el IC, y menor la precisión obtenida.

## Cálculo del tamaño de la muestra

Para calcular el tamaño de la muestra debe conocerse:

- La *variabilidad* del parámetro que se desea estimar. Si no se conoce, puede obtenerse una aproximación a partir de datos propios o de otras investigaciones, o un estudio piloto. En el caso de las variables cuantitativas se mide por la variancia, y en el de las cualitativas, por el producto  $P \cdot (1 - P)$ .
- La *precisión* con que se desea obtener la estimación, es decir, la amplitud del IC. Cuanto más precisa se desee, más estrecho deberá ser este intervalo, y más sujetos deberán ser estudiados. La precisión debe fijarse previamente, en función de la finalidad de la estimación. En algunos casos puede requerirse una gran precisión, mientras que en otros, si sólo se necesita conocer aproximadamente entre qué valores se encuentra el parámetro, se requerirá una menor precisión y, consecuentemente, menos sujetos.
- El *nivel de confianza* deseado, que habitualmente se fija en el 95%, correspondiente a un valor  $\alpha$  de 0,05. Indica el grado de confianza que se tiene de que el verdadero valor del parámetro en la población se sitúa en el intervalo obtenido. Cuanta más confianza se desee, menor será el valor de  $\alpha$ , y más elevado el número de sujetos necesario.

De estos tres elementos, sólo debe conocerse la variabilidad del parámetro, ya que tanto la precisión como el nivel de confianza se fijan a partir de los intereses del investigador.

### Estimación de una proporción

La fórmula para el cálculo del número de sujetos para estimar una proporción se presenta en el cuadro 15.1. También puede utilizarse la tabla A que se encuentra al final del libro.

**Ejemplo 15.1.** Un estudio se propone estimar el porcentaje de pacientes hipertensos de un centro que están controlados. A partir de datos previos, se estima que debe estar situado alrededor del 40% ( $p = 0,40$ ). Se quiere realizar la estimación con una precisión de  $\pm 4\%$  ( $i = 0,04$ ) y una confianza del 95% ( $1 - \alpha = 0,95$ ;  $\alpha = 0,05$ ;  $Z\alpha = 1,96$ ). Aplicando

**Cuadro 15.1 Fórmulas para el cálculo del número de sujetos necesarios para la realización de un estudio cuyo objetivo es la estimación de una media o una proporción**

#### Estimación de una proporción (variable cualitativa)

$$N = \frac{Z\alpha^2 P(1-P)}{i^2}$$

#### Estimación de una media (variable cuantitativa)

$$N = \frac{Z\alpha^2 s^2}{i^2}$$

N: número de sujetos necesarios.

Z $\alpha$ : valor de Z correspondiente al riesgo  $\alpha$  fijado (tabla 15.4).

i: valor de la proporción que se supone existe en la población.

s<sup>2</sup>: variancia de la distribución de la variable cuantitativa que se supone que existe en la población.

i: precisión con que se desea estimar el parámetro (2·i es la amplitud del intervalo de confianza).

la fórmula correspondiente, se calcula que se precisan 576 sujetos. La tabla A proporciona el mismo resultado.

Esta estimación del número de sujetos realizada a priori no es más que una aproximación. Si en el ejemplo el porcentaje real de pacientes controlados es diferente del 40%, puede alterarse el grado de precisión de la estimación.

En la tabla 15.1 puede apreciarse cómo diferentes grados de precisión y diferentes valores del porcentaje esperado modifican el número de sujetos necesarios. Así, por ejemplo, para estimar el porcentaje del 40% del ejemplo 15.1 con el mismo nivel de confianza del 95%, los 576 sujetos se convierten en 9.220 cuando se desea una precisión muy alta ( $i = 0,01$ ), o en tan sólo 92 si se es menos exigente ( $i = 0,1$ ). Modificando cualquiera de estos valores, puede obtenerse un número de individuos que se aproxime al «deseado» o al disponible. Debe evitarse cualquier manipulación del cálculo, ya que, al reducir el número de sujetos que se van a estudiar, también disminuye el grado de precisión con que el parámetro va a ser estimado y aumenta la amplitud del IC.

En el cálculo del tamaño de la muestra también debe tenerse en cuenta la estrategia de análisis y cómo se presentarán los resultados. Así, si en el

**Tabla 15.1 Influencia de la precisión de la estimación y del valor supuesto de la proporción que se desea estimar sobre el número de sujetos necesarios**

<b>Influencia de la precisión</b>		
<b>P</b>	<b>i</b>	<b>N</b>
0,40	0,100	92
0,40	0,075	164
0,40	0,050	369
0,40	0,025	1.475
0,40	0,010	9.220
<b>Influencia de la proporción</b>		
<b>P</b>	<b>i</b>	<b>N</b>
0,10	0,05	138
0,20	0,05	246
0,30	0,05	323
0,40	0,05	369
0,50	0,05	384

Todos los cálculos se han realizado con el grado de confianza del 95%. P: valor supuesto de la proporción que se desea estimar. i: medida de la precisión de la estimación, que corresponde a la mitad del intervalo de confianza. N: número de sujetos necesarios para realizar la estimación de P con la precisión deseada y el 95% de confianza.

en el ejemplo 15.1 los investigadores presentan el resultado principal en función del sexo, la estratificación hará que la estimación se haya obtenido en un número menor de sujetos, por lo que la precisión será menor de la deseada.

En algunas ocasiones no se conoce el valor aproximado del parámetro que se está buscando. Si no existen datos de la literatura que resulten útiles, o si no puede realizarse una prueba piloto para obtener una primera aproximación a dicho valor, puede adoptarse la postura de la máxima indeterminación. Esta actitud consiste en suponer que el porcentaje que se quiere estimar se sitúa alrededor del 50%, ya que es el valor que requiere una mayor cantidad de individuos para una precisión determinada, de forma que el número obtenido supera cualquier otra combinación de parámetros. Esta actitud también es útil, por ejemplo, cuando se realizan encuestas en las que interesa estimar múltiples parámetros. Su inconveniente es que requiere el estudio de un número de sujetos a veces muy superior al realmente necesario.

## Estimación de una media

El cálculo es similar al de la situación anterior (ver cuadro 15.1). También deben fijarse el nivel de confianza y la precisión de la estimación, que se traducen en la amplitud del intervalo alrededor de la media que se desea estimar. En las variables cuantitativas, la medida de la variabilidad viene proporcionada por la variancia de su distribución en la población. La tabla B del final del libro proporciona una fácil solución alternativa al cálculo matemático.

**Ejemplo 15.2.** Supongamos que se desea estimar la tensión arterial diastólica (TAD) de los pacientes diabéticos registrados en un centro de salud. Por estudios previos se conoce que la desviación estándar de la TAD en sujetos diabéticos es de 25 mmHg ( $s = 25 \text{ mmHg}$ ;  $s^2 = 625 \text{ mmHg}^2$ ). Se quiere realizar la estimación con una confianza del 95% ( $1 - \alpha = 0,95$ ;  $\alpha = 0,05$ ) y una precisión de 5 mmHg ( $i = 5$ ). Aplicando la fórmula, se puede determinar que son necesarios 96 sujetos. Esta cifra debe aumentarse en función del porcentaje de sujetos en que se prevea que no podrá determinarse la TAD. El mismo resultado se obtiene consultando la tabla B en la línea correspondiente al valor  $i/s = 5/25 = 0,200$ .

Si no se dispone de una estimación de la variabilidad, puede utilizarse una regla práctica, que consiste en determinar la diferencia entre los valores máximo y mínimo esperables. Dado que se asume que esta variable es de distribución normal, el intervalo  $m \pm (2s)$ , siendo  $m$  la media y  $s$  la desviación estándar de la distribución, incluye aproximadamente el 95% de los valores posibles, por lo que, al dividir dicha amplitud de valores entre 4, puede obtenerse una cierta aproximación al valor de la desviación estándar  $s$ .

## Corrección para poblaciones finitas

En los cálculos anteriores no ha intervenido el tamaño de la población, ya que en ellos se ha asumido que es infinito. Sin embargo, en muchas ocasiones es preciso obtener una muestra de una población de tamaño conocido (finito). En esta situación puede aplicarse la siguiente fórmula, que ajusta el número de individuos necesarios según el tamaño de la población:

$$n_a = n[1 + (n/N)]$$

donde  $n_a$  es el número de sujetos necesarios,  $n$  es el número de sujetos calculado para poblaciones infinitas y  $N$  es el tamaño de la población.

**Ejemplo 15.3.** En el ejemplo 15.1 se ha calculado que son necesarios 576 sujetos para estimar el

porcentaje de hipertensos controlados de un centro, realizando determinadas asunciones. Si la población de referencia es de 1.000 hipertensos, aplicando la fórmula anterior podría determinarse que son necesarios 365 pacientes hipertensos.

## ESTUDIOS QUE COMPARAN DOS GRUPOS

Muchos estudios tienen por objetivo responder a preguntas del siguiente tipo: ¿es más eficaz el fármaco A que el B en el tratamiento de una enfermedad? ¿Existe un aumento del riesgo de malformaciones congénitas en los niños nacidos de madres que han padecido cierto trastorno durante el embarazo?, etc.

### Consideraciones previas: pruebas de contraste de hipótesis

Bajo las preguntas que se han planteado en el párrafo anterior subyace una hipótesis, por ejemplo, la de que B es más eficaz que A en el tratamiento de los hipertensos esenciales de edad avanzada. En la realidad, esta hipótesis puede ser verdadera o falsa, y eso es lo que se trata de averiguar. Dado que el investigador no puede estudiar todos los hipertensos esenciales de edad avanzada, la hipótesis debe contrastarse en una muestra de esta población. A partir únicamente de los datos recogidos en un estudio, no puede aceptarse o rechazarse absolutamente la hipótesis que se contrasta. Existe cierto riesgo de error, sea cual sea la conclusión a la que se llegue.

#### Error $\alpha$ y error $\beta$

Desde el punto de vista estadístico, el investigador se enfrenta al problema del contraste de una

hipótesis mediante la asunción de que no existen diferencias en la eficacia de los dos fármacos A y B (*hipótesis nula*). A partir de los resultados observados en la muestra, el investigador utilizará las pruebas de significación estadística para evaluar si existe la suficiente evidencia que le permita rechazar esta hipótesis nula y, consecuentemente, aceptar la hipótesis alternativa de que sí existen diferencias entre ambas terapéuticas.

El *error tipo I*, conocido también como *error  $\alpha$*  (ver tabla 15.2), se comete cuando el investigador rechaza la hipótesis nula, siendo ésta verdadera en la población. Es equivalente a encontrar un resultado falso positivo, ya que el investigador concluye que hay una diferencia, cuando en realidad no existe.

El *error tipo II*, o *error  $\beta$*  (ver tabla 15.2), se comete en la situación contraria, cuando el investigador no rechaza la hipótesis nula, siendo ésta falsa en la población. Es equivalente a un resultado falso negativo, ya que el investigador concluye que ha sido incapaz de encontrar una diferencia que existe en la realidad. Si  $\beta$  representa la probabilidad de un resultado falso negativo, su complementario  $1 - \beta$ , conocido como *poder* o *potencia estadística*, representa la probabilidad de observar en la muestra una determinada diferencia o un efecto, si existen en la población.

Para calcular el tamaño de la muestra, se fijan a priori los riesgos que se está dispuesto a asumir de cometer estos errores. Lo ideal sería poder reducirlos a cero, pero ello no es posible. Una de las formas para disminuirlos consiste en aumentar el número de sujetos estudiados.

### Hipótesis unilaterales e hipótesis bilaterales

Desde el punto de vista estadístico, otro aspecto que afecta al cálculo del tamaño de la muestra es el de la formulación precisa de la hipótesis que se

**Tabla 15.2 Tipos de error aleatorio en una prueba estadística de contraste de hipótesis**

	Realidad (población)		Error tipo I $\alpha$
	Existe diferencia o asociación ( $H_0$ falsa)	No existe diferencia o asociación ( $H_0$ cierta)	
Resultado de la prueba (muestra)	Diferencia o asociación significativa (rechazo de $H_0$ )	No error	Error tipo I $\alpha$
	Diferencia o asociación no significativa (no rechazo de $H_0$ )	Error tipo II $\beta$	
Ho: hipótesis nula.			

contrasta. Si se pretende determinar si dos fármacos A y B difieren en su eficacia, sin decir cuál de ellos ha de ser mejor o peor, se habla de *hipótesis bilateral*, o de dos colas (fig. 15.1). En cambio, si la hipótesis alternativa es que el fármaco A es más eficaz que el B, se habla de *hipótesis unilateral*, o de una cola. En este caso, interesa únicamente evaluar si B es, como mínimo, tan eficaz como A, es decir, tan sólo una de las direcciones de la comparación.

Desde un punto de vista estricto, son preferibles las hipótesis bilaterales, ya que son más conservadoras y requieren un mayor número de participantes.

### Magnitud de la diferencia o de la asociación

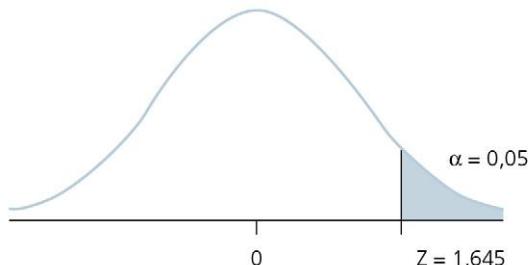
La magnitud de la diferencia o de la asociación tiene una gran influencia sobre el tamaño de la muestra. Si la diferencia es muy amplia (p. ej., si la eficacia del fármaco A es el doble que la del B), se podrá detectar con más facilidad que si es de escasa magnitud, y requerirá un número menor de sujetos. En la tabla 15.3 puede apreciarse cómo una diferencia de 40 mg/100 ml en la media de colesterol

de dos muestras puede ser significativa, aunque se hayan estudiado tan sólo 20 sujetos, mientras que una pequeña diferencia de 3 mg/100 ml no alcanza suficiente significación, aunque se hayan estudiado 200 individuos en cada grupo.

Para calcular el tamaño de la muestra, debe definirse la mínima magnitud de la diferencia o de la asociación que se desea detectar y que se considera de relevancia clínica. Esta diferencia debe fijarse en términos realistas, dentro del rango de valores esperables. A menudo, es tentador utilizar diferencias más amplias, ya que se obtiene un tamaño de la muestra más reducido. Esta actitud debe evitarse, ya que implica que diferencias menores a la establecida, a pesar de que pueden ser de importancia clínica, no alcanzarán el grado de significación estadística con el número de sujetos estudiados. Si, por ejemplo, se utiliza una diferencia del 50% en la eficacia de los dos fármacos en el cálculo del tamaño de la muestra, se necesitará un número reducido de sujetos, pero una diferencia del 40%, que puede ser muy importante desde el punto de vista clínico, no será estadísticamente significativa, ya que el investigador ha decidido considerar como importantes sólo las diferencias superiores al 50%.

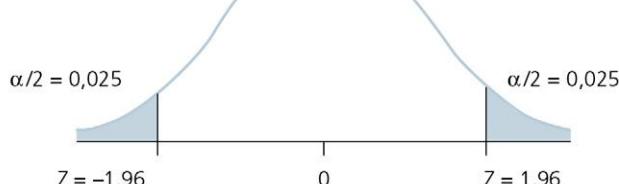
#### Prueba unilateral\*

$$\begin{aligned} H_0: & A \leq B \\ H_a: & A > B \end{aligned}$$



#### Prueba bilateral\*\*

$$\begin{aligned} H_0: & A = B \\ H_a: & A \neq B \end{aligned}$$



\* Especifica la dirección de la diferencia.

\*\* Interesa determinar si A y B son diferentes, sin especificar la dirección de la diferencia. Dado que se asume que la distribución es normal, el valor de α se distribuye simétricamente en las dos colas de la distribución.

$H_0$ : hipótesis nula.

$H_a$ : hipótesis alternativa.

**Figura 15.1** Valores de la distribución normal tipificada (Z) correspondientes al valor  $\alpha$  del 5% en las pruebas estadísticas unilaterales y bilaterales.

**Tabla 15.3 Influencia del número de sujetos estudiados y la magnitud de la diferencia sobre la significación estadística**

**Ejemplo 1. Una diferencia de gran magnitud es estadísticamente significativa incluso con un reducido número de sujetos**

	n	Colesterolemia (media ± DE)	t	p
Muestra 1	20	260 ± 20		
			6,32	< 0,001
Muestra 2	20	220 ± 20		

**Ejemplo 2. Una diferencia de escasa magnitud no es estadísticamente significativa, aunque se haya estudiado un elevado número de sujetos**

	n	Colesterolemia (media ± DE)	t	p
Muestra 1	200	218 ± 20		
			1,5	0,14
Muestra 2	200	215 ± 20		

n: número de sujetos de la muestra. DE: desviación estándar. t: valor de la t de Student-Fisher de comparación de dos medidas independientes. p: probabilidad de cometer un error  $\alpha$ . Habitualmente, valores menores de 0,05 se aceptan como evidencia suficiente para decidir que la diferencia es estadísticamente significativa.

En función del enfoque del estudio, esta magnitud puede fijarse como una diferencia entre grupos, o bien, en términos relativos, como una relación de riesgos.

### Cálculo del tamaño de la muestra

Para calcular el tamaño de la muestra deben utilizarse los siguientes elementos principales (cuadro 15.2):

- Definir la hipótesis que se va a contrastar, precisando si es unilateral o bilateral.
- Establecer el riesgo de cometer un error de tipo I ( $\alpha$ ) que se está dispuesto a aceptar. Habitualmente suele aceptarse un riesgo  $\alpha$  del 5%, y preferiblemente con hipótesis bilaterales, ya que son más conservadoras.
- Establecer, así mismo, el riesgo que se acepta de cometer un error tipo II ( $\beta$ ). Habitualmente se sitúa entre el 5 y el 20%. La elección variará en función de las consecuencias que pueda tener cometer dicho error. A menudo es más fácil enfrentar esta decisión a partir del concepto de poder o potencia estadística ( $1 - \beta$ ), que es la capacidad del estudio para detectar una determinada diferencia. Aceptar el riesgo de cometer un error  $\beta$  del 20% significa que, si la diferencia que se busca existe en la realidad, el estudio tiene un 80% de probabilidades de detectarla.
- Definir la mínima magnitud de la diferencia, el efecto o la asociación que se quiera ser capaz

### Cuadro 15.2 Elementos que intervienen en el cálculo del número de sujetos necesarios para la realización de un contraste de hipótesis

- 1 Hipótesis uni o bilateral
- 2 Riesgo aceptado de cometer un error  $\alpha$
- 3 Riesgo aceptado de cometer un error  $\beta$  (potencia:  $1 - \beta$ )
- 4 Magnitud mínima de la diferencia o asociación que se considera importante detectar
- 5 Variabilidad de la variable de respuesta en el grupo de referencia
- 6 Otros factores:
  - Prueba estadística que se utilizará en el análisis
  - Tipos especiales de diseño del estudio
  - Comparación de más de dos grupos
  - Grupos de tamaño diferente
  - Equivalencia de las intervenciones
  - Estratificación por las características basales
  - Múltiples variables de respuesta
  - Porcentaje de pérdidas y abandonos esperado durante el seguimiento

de detectar. Debe basarse en datos de estudios previos o de la literatura que definan el rango de los valores esperables, y en la mínima magnitud que se considere de relevancia clínica.

- Es necesario, también, disponer de alguna medida de la variabilidad de la variable de respuesta en la población o el grupo de referencia.

De estos cinco elementos, sólo debe conocerse el último, ya que los cuatro primeros son fijados por el investigador. A continuación, se aplica la fórmula correspondiente, que depende de la prueba estadística que vaya a ser utilizada en el análisis. En el cuadro 15.3 se presentan las fórmulas de uso más habitual, correspondientes a la comparación de dos variables dicotómicas (dos proporciones) o de dos variables cuantitativas (dos medias). Las tablas 15.4 y 15.5 recogen los valores de la variable normal tipificada, correspondientes a los riesgos  $\alpha$  y  $\beta$  utilizados con mayor frecuencia en dicho cálculo, respectivamente.

Las fórmulas pueden resultar algo complicadas, por lo que se recomienda utilizar tablas como las presentadas al final del libro (tablas C y D).

**Cuadro 15.3 Fórmulas para el cálculo del número de sujetos necesarios por grupo en un estudio cuyo objetivo es la comparación de dos muestras del mismo tamaño**

Comparación de dos proporciones (variable cualitativa)

$$N = \frac{[Z\alpha\sqrt{[2 \cdot P \cdot (1-P)]} + Z\beta\sqrt{[P_1 \cdot (1-P_1) + P_2 \cdot (1-P_2)]}]^2}{(P_1 - P_2)^2}$$

Comparación de dos medias (variable cuantitativa)

$$N = \frac{2 \cdot (Z\alpha + Z\beta)^2 \cdot s^2}{d^2}$$

N: número de sujetos necesarios en cada uno de los grupos.

Z $\alpha$ : valor de Z correspondiente al riesgo  $\alpha$  fijado (tabla 15.4).

Z $\beta$ : valor de Z correspondiente al riesgo  $\beta$  fijado (tabla 15.5).

P<sub>1</sub>: valor de la proporción que se supone existe en el grupo de referencia.

P<sub>2</sub>: valor de la proporción que se supone existe en el grupo de estudio.

P<sub>2</sub> – P<sub>1</sub>: valor mínimo de la diferencia que se desea detectar (variable cualitativa).

P: media ponderada de las proporciones P<sub>1</sub> y P<sub>2</sub>.

s<sup>2</sup>: variancia de la distribución de la variable cuantitativa que se supone que existe en el grupo de referencia.

d<sup>2</sup>: valor mínimo de la diferencia que se desea detectar (variable cuantitativa).

## Comparación de dos proporciones

La comparación de dos proporciones es, probablemente, la prueba de contraste de hipótesis más utilizada en investigación clínica. Se trata de estudios que comparan dos grupos, y para ello utilizan una variable de respuesta cualitativa dicotómica.

**Ejemplo 15.4.** Un estudio tiene por objetivo determinar si un nuevo tratamiento T consigue un mayor porcentaje de éxitos en las sobreinfecciones respiratorias que el tratamiento estándar E. Lo primero que debe conocerse es el porcentaje de curaciones en pacientes de características similares a los que van a ser estudiados, obtenido con el tratamiento estándar E. Supongamos que esta cifra se sitúa alrededor del 40% (P<sub>1</sub> = 0,4).

El siguiente paso es determinar la diferencia mínima que se quiere detectar, es decir, responder a la pregunta: ¿a partir de qué porcentaje de éxitos con el nuevo tratamiento se considerará que éste es mejor que E y, por lo tanto, se estará dispuesto a modificar la pauta terapéutica habitual? Es decir, si el porcentaje de individuos curados con T es del 41%, ¿puede considerarse que esta diferencia del 1% es un resultado lo suficientemente importante para modificar la pauta terapéutica? ¿O se exigirá un mínimo, por ejemplo, del 50% de éxitos? La respuesta a estas preguntas depende de muchos factores, tales como la seguridad del fármaco, la facilidad de administración o el coste, entre otros. Supongamos que los investigadores consideran que, si se cura el 50% de pacientes con T (P<sub>2</sub> = 0,5), se aceptará como la elección terapéutica.

A continuación, sólo falta determinar los niveles de riesgo de cometer algún tipo de error aleatorio que se está dispuesto a asumir. Supongamos que se acepta el nivel de riesgo  $\alpha$  habitual del 5% con una hipótesis bilateral y un riesgo  $\beta$  del 20% (potencia: 1 –  $\beta$  = 0,80).

Aplicando la fórmula del cuadro 15.3, puede calcularse que son necesarios 387 sujetos por grupo de estudio. Con la tabla C del final del libro se obtiene la misma cifra.

Este número indica los sujetos que deben finalizar el estudio para tener un 80% de probabilidades de detectar una diferencia igual o superior a la fijada, con un nivel de error  $\alpha$  del 5%. Por lo tanto, hay que incrementarlo en función del número de pérdidas de seguimiento y de abandonos que se prevea que ocurrirán durante el estudio.

El tamaño de la muestra requerido es mayor cuanto más pequeña es la diferencia que se desea detectar y menos elevados son los riesgos  $\alpha$  y  $\beta$  que

**Tabla 15.4** Valores de  $Z_\alpha$  utilizados con mayor frecuencia en el cálculo del tamaño de la muestra

$\alpha$	$Z_\alpha$	
	Prueba unilateral	Prueba bilateral
0,200	0,842	1,282
0,150	1,036	1,440
0,100	1,282	1,645
0,050	1,645	1,960
0,025	1,960	2,240
0,010	2,326	2,576

**Tabla 15.5** Valores de  $Z_\beta$  utilizados con mayor frecuencia en el cálculo del tamaño de la muestra

Potencia		
$\beta$	$(1 - \beta)$	$Z_\beta$
0,01	0,99	2,326
0,05	0,95	1,645
0,10	0,90	1,282
0,15	0,85	1,036
0,20	0,80	0,842
0,25	0,75	0,674
0,30	0,70	0,524
0,35	0,65	0,385
0,40	0,60	0,253
0,45	0,55	0,126
0,50	0,50	0,000

se aceptan. Es importante destacar que una misma diferencia del 5% entre dos porcentajes requiere el máximo número de pacientes cuando éstos se sitúan alrededor del 50%. Ello se debe a que, si el porcentaje del grupo de referencia es del 5%, una diferencia del 5% implica que el grupo de estudio debe pasar a un 10%, es decir, dobla la eficacia. En cambio, si el porcentaje del grupo de referencia es del 50%, una diferencia del 5% implica solamente un incremento relativo del 10%, es decir, pasar del 50 al 55%.

### Estimación de un riesgo relativo

Si el objetivo del estudio es determinar la magnitud de la asociación en términos relativos, la medida que se utilizará en los estudios prospectivos (ensayos

clínicos y estudios de cohortes) es el riesgo relativo (RR). Lo que debe fijarse en este caso es la mínima magnitud del RR que se quiere ser capaz de detectar. En esta situación puede utilizarse la misma fórmula que en el caso de la comparación de dos proporciones, teniendo en cuenta que  $RR = P_2/P_1$ . La magnitud de la asociación que se quiere detectar corresponde al mínimo RR que se considerará de importancia clínica, donde  $P_1$  es el riesgo de desarrollar la enfermedad en el grupo no expuesto y  $P_2$  el riesgo de desarrollarla en el grupo expuesto ( $P_2 = P_1 \cdot RR$ ).

La tabla D del final del libro presenta el resultado del cálculo del número de sujetos necesarios para los estudios que estiman un RR en las situaciones más habituales.

**Ejemplo 15.5.** Se pretende realizar un estudio de cohortes con el objetivo de estimar el riesgo de desarrollar infarto agudo de miocardio (IAM) asociado a la hipercolesterolemia. El riesgo esperado en la cohorte no expuesta ( $P_1$ ) es del 10% a los 10 años (incidencia acumulada). El riesgo relativo (RR) que se considera relevante detectar es de 2. El valor del riesgo de cometer un error  $\alpha$  será el convencional del 5% (hipótesis bilateral) y el de cometer un error  $\beta$ , del 20% (potencia:  $1 - \beta = 0,80$ ). A partir de estos datos, puede calcularse el valor de  $P_2$ :

$$P_2 = P_1 \cdot RR = 0,10 \cdot 2 = 0,20 \text{ (o } 20\%)$$

Aplicando la fórmula del cuadro 15.3 o la tabla D del final del libro, puede determinarse que son necesarios 199 sujetos por grupo. Este número representa el tamaño mínimo tanto del grupo de expuestos como del de no expuestos. Por lo tanto, si desea estudiar una cohorte general, o si se deben identificar los sujetos expuestos a partir de una muestra de la población general, se deberá determinar el número de sujetos que habrá que examinar para identificar

el número suficiente de expuestos. Si la frecuencia de la exposición en la población general ( $f$ ) fuera de un 20%, sería necesario examinar a  $N/f = 199/0,2 = 995$  sujetos para encontrar a 199 expuestos.

### Estimación de una *odds ratio*

En los estudios de casos y controles, la magnitud de la asociación se estima mediante la *odds ratio* (OR). Se trata de una situación similar a la de la estimación de un RR, en la que debe conocerse la proporción de exposición esperada en los controles ( $P_1$ ) y fijar la magnitud de la OR que se desea detectar. Análogamente a la situación anterior, la proporción esperada de exposición entre los casos es:

$$P_2 = P_1 \cdot OR [1 + P_1 \cdot (OR - 1)]$$

La tabla E del final del libro presenta el resultado del cálculo del número de sujetos necesarios para los estudios que estiman una OR en las situaciones más habituales.

**Ejemplo 15.6.** Un estudio de casos y controles pretende evaluar el riesgo de padecer un IAM asociado al uso de anticonceptivos orales (AO) en mujeres jóvenes. Se estima que la proporción esperada de uso de AO en la población candidata a ser incluida en el grupo control es del 30% ( $P_1 = 0,30$ ). El incremento mínimo del riesgo que se quiere detectar es de 3 ( $OR = 3$ ). El riesgo  $\alpha$  es de 0,05 (hipótesis bilateral) y el riesgo  $\beta$  es de 0,10 (potencia:  $1 - \beta = 0,90$ ). Por lo tanto:

$$P_2 = (0,3 \cdot 3) / [1 + 0,3 \cdot (3 - 1)] = 0,56$$

Aplicando la fórmula del cuadro 15.3, puede determinarse que son necesarios 74 sujetos por grupo. La tabla E del final del libro proporciona un resultado similar.

### Comparación de dos medias

Para calcular el tamaño de la muestra debe conocerse el valor de la variancia de la distribución de la variable de respuesta en el grupo de referencia, decidir la mínima diferencia que se considera de relevancia clínica y si interesa poder detectar y fijar los niveles de error  $\alpha$  y  $\beta$  que se aceptan.

**Ejemplo 15.7.** Un estudio se propone comparar los niveles de fructosamina de dos grupos de diabéticos sometidos a dos pautas terapéuticas diferentes, T1 y T2, con la finalidad de determinar cuál de ellas consigue un mayor control de la enfermedad. Se utiliza habitualmente la pauta T1, con la que se consigue unos valores medios de fructosamina de 3 mmol/l, con una desviación estándar de 0,625 mmol/l. Por lo

tanto, el valor de la variancia  $s^2$  es de  $(0,625)^2 = 0,39$ . A continuación, se establece que la diferencia mínima entre ambos grupos que se considera de relevancia clínica es de 0,5 mmol/l ( $d = 0,5$ ). El riesgo de error  $\alpha$  que se está dispuesto a asumir es de 0,05. Dado que se pretende evaluar qué pauta es más eficaz, se trata de una hipótesis bilateral. Interesa tener una capacidad del 90% para detectar la diferencia, si existe realmente ( $1 - \beta = 0,90$ ).

Para utilizar la tabla F del final del libro debe calcularse el valor  $d/s$ , que corresponde al cociente entre la diferencia que se quiere detectar y la desviación estándar de la variable de respuesta en el grupo de referencia. En este caso, corresponde a  $0,5/0,625 = 0,8$ . Buscando este valor en la primera columna de las tablas, y leyendo la fila correspondiente al valor de  $\alpha$  de 0,05 en una hipótesis bilateral, en el punto de intersección con la potencia de 0,90 se obtiene un tamaño de 33 sujetos por grupo; es decir, en total, debe incluirse en el estudio un mínimo de 66 personas. El mismo resultado se obtiene utilizando la fórmula del cuadro 15.3.

### Comparación de grupos cuando la variable de respuesta es ordinal

Cuando la variable de respuesta se mide en una escala ordinal, o cuando en el análisis deben utilizarse pruebas estadísticas no paramétricas, el cálculo es más complejo. Por criterios de operatividad, suele calcularse el tamaño como si fuera a utilizarse una prueba paramétrica. La estimación será tanto mejor cuanto menos se aleje la distribución de la variable de la ley normal.

En el anexo 7 se presenta el cálculo del tamaño de la muestra en otras situaciones.

### Equivalencia de dos intervenciones

En ocasiones se diseña un estudio para determinar si una nueva intervención, que ofrece alguna ventaja sobre la de comparación, como ser menos costosa o más segura, es tan eficaz como ella (estudios de equivalencia) o no es terapéuticamente inferior (estudios de no-inferioridad). En esta situación, la determinación del tamaño de la muestra requiere tener en cuenta algunas consideraciones especiales. El problema radica en que no existe un procedimiento estadístico para demostrar la equivalencia de dos intervenciones. La ausencia de una diferencia estadísticamente significativa entre ellas no debe interpretarse como prueba de que sean terapéuticamente equivalentes.

La demostración de la equivalencia requiere descartar pequeñas diferencias en la respuesta observada. Para considerar que dos intervenciones son

equivalentes, la diferencia entre ellas debe ser tan pequeña que pueda considerarse clínicamente irrelevante. Por ello, debe definirse cuál es la máxima diferencia que puede existir entre dos intervenciones para considerarlas equivalentes. Este margen de equivalencia, conocido como *delta*, debe ser menor que la mínima diferencia clínicamente importante, que es la que se utilizaría para determinar el tamaño de la muestra en un estudio de superioridad. En los estudios de equivalencia, una forma de calcular el tamaño de la muestra necesario es tomar el valor del margen de equivalencia como la diferencia que se desea ser capaz de detectar. Sin embargo, este procedimiento infraestima el número de sujetos necesarios.

Tal como se comenta en el capítulo 5, la interpretación de los resultados de los estudios de equivalencia se basa en el análisis del intervalo de confianza (IC) de la diferencia observada, de manera que se considera que existe equivalencia si todo el IC se sitúa dentro de los márgenes de equivalencia (o si su límite inferior está por encima de dicho margen en los estudios de no-inferioridad). Para obtener un IC lo suficientemente estrecho es necesario estudiar un número de sujetos sustancialmente más elevado que en los estudios de superioridad. Por lo tanto, en el cálculo del tamaño de la muestra en este tipo de estudios debe tenerse en cuenta no solamente el margen de equivalencia establecido, sino también el valor esperado de la respuesta (Tamayo-Sarver et al, 2005), tal como se muestra en la fórmula del cuadro 15.4. La tabla G del final del libro presenta el número de sujetos necesarios para estudios de equivalencia en las situaciones más habituales cuando la variable de respuesta es dicotómica.

**Ejemplo 15.8.** Un estudio se propone evaluar la eficacia de una pauta monodosis para el tratamiento de las infecciones urinarias frente a la pauta habitual de 7 días, cuya eficacia se sitúa alrededor del 90% ( $P=0,90$ ). La monodosis es menos costosa y facilita el cumplimiento del paciente, por lo que interesa determinar si la monodosis y la pauta de 7-10 días pueden considerarse equivalentes. Los investigadores definen que considerarán que lo son si la diferencia de eficacia entre ellas no supera el 10% (margen de equivalencia  $d=0,10$ ). Por otro lado, los investigadores esperan encontrar que la monodosis es ligeramente inferior, en un 5%, a la pauta de 7 días ( $d^*=0,05$ ). Se aceptan los niveles de error  $\alpha$  del 0,05 (hipótesis bilateral) y  $\beta$  del 20% (potencia:  $1-\beta=0,80$ ). Aplicando la fórmula del cuadro 15.4 para una variable dicotómica, puede determinarse que son necesarios 565 sujetos por grupo. La tabla G del final del libro proporciona el mismo resultado.

#### Cuadro 15.4 Fórmula para el cálculo del tamaño de la muestra en un estudio de equivalencia

Equivalencia entre dos proporciones (variable cualitativa)

$$N = \frac{2 \cdot P \cdot (1-P) \cdot (Z\alpha + Z\beta)^2}{(d - d^*)^2}$$

Equivalencia entre dos medias (variable cuantitativa)

$$N = \frac{2 \cdot s^2 \cdot (Z\alpha + Z\beta)^2}{(d - d^*)^2}$$

N: número de sujetos necesarios en cada uno de los grupos.

Z $\alpha$ : valor de Z correspondiente al riesgo  $\alpha$  fijado (tabla 15.4).

Z $\beta$ : valor de Z correspondiente al riesgo  $\beta$  fijado (tabla 15.5).

P: proporción que se espera en el grupo control (en caso de una variable cualitativa)

S $^2$ : variancia de la distribución de la variable de respuesta (en caso de una variable cuantitativa)

d: diferencia máxima entre las respuestas a ambos tratamientos que se tolerará para concluir que son equivalentes (margen de equivalencia)

d $^*$ : diferencia esperada entre las respuestas a ambos tratamientos

El mismo resultado se obtendría si el estudio se plantease con una hipótesis de no-inferioridad, en el que los investigadores definieran que considerarían que la monodosis es mejor alternativa terapéutica si no es más de un 10% inferior a la pauta de 7 días.

Este procedimiento suele sobreestimar el número de sujetos necesarios. Para obtener un cálculo más ajustado, puede utilizarse una fórmula basada en el cálculo del IC, tal como se comenta a continuación.

#### CÁLCULO DEL TAMAÑO DE LA MUESTRA BASADO EN INTERVALOS DE CONFIANZA

Algunos autores piensan que si los resultados de un estudio se analizan, se interpretan y se presentan mediante IC, el cálculo del tamaño de la muestra debería basarse también en IC.

Al interpretar los resultados de un estudio se recomienda tener en cuenta la diferencia mínima que se considera relevante desde el punto de vista clínico (ver capítulo 27). Si el límite inferior del IC se encuentra por encima del valor de esta diferencia mínima, se considera que el resultado del estudio es no solamente estadísticamente significativo, sino también concluyente

a favor de la existencia de una diferencia clínicamente relevante. De modo similar, si el valor superior del IC está por debajo del valor de la diferencia mínima, el resultado es concluyente a favor de la ausencia de una diferencia clínicamente relevante.

En algunas ocasiones existe desacuerdo entre el método del cálculo del tamaño de la muestra basado en el contraste estadístico de hipótesis y la interpretación de los resultados basada en los IC. Este desacuerdo es evidente cuando el investigador ha calculado el tamaño de la muestra necesario para detectar, con una determinada potencia estadística, una diferencia mínima entre dos tratamientos, pero el análisis del IC del resultado muestra que valores inferiores a dicha diferencia son compatibles con los datos observados. En esta situación, el investigador no puede concluir con certeza cuál es la mejor alternativa terapéutica.

**Ejemplo 15.9.** En un estudio que comparaba la cirugía convencional (P1) con la radical (P2) en el tratamiento del cáncer gástrico, el cálculo del tamaño de la muestra se basó en la asunción de una diferencia en la supervivencia a los 5 años del 14% ( $P_1 = 20\%$ ;  $P_2 = 34\%$ ) (Fayers et al, 2000). Tomando como base estas tasas de supervivencia, se requerían 400 pacientes por grupo si se deseaba una potencia estadística del 90%. Al inicio del estudio, un grupo de 26 cirujanos llenaron un cuestionario en el que manifestaban que una mejora en la tasa de supervivencia del 9,5% podía considerarse clínicamente relevante. Si los resultados del estudio hubieran sido los de las asunciones usadas para el cálculo del tamaño de la muestra, y 400 pacientes hubieran finalizado el estudio en cada grupo, se obtendría una diferencia del 14% con un IC 95% cuyos límites serían 5,4% y 22,6%. Este resultado, aunque es estadísticamente significativo, no descarta que la diferencia en la supervivencia pueda ser inferior al 9,5%, por lo que no sería concluyente.

Este problema se puede evitar basando el cálculo del tamaño de la muestra en los IC mediante la siguiente fórmula:

$$n = \frac{8Z_{\alpha}^2 p(1-p)}{W_{IC}^2}$$

donde  $n$  es el número de individuos por grupo,  $Z_{\alpha}$  es el valor de  $Z$  correspondiente al riesgo  $\alpha$  fijado,  $P$  es la proporción del parámetro estimado en la población ( $P = [P_1 + P_2]/2$ ) y  $W_{IC}$  es la amplitud del IC.

**Ejemplo 15.10.** Consideraremos el estudio planteado en el ejemplo anterior (Fayers et al, 2000). Aplicando la fórmula:

$$n = \frac{8 \cdot 1,96^2 \cdot 0,27 \cdot 0,73}{W_{IC}^2}$$

Si pensamos que la diferencia más realista entre grupos es del 14% ( $P_1 = 20\%$ ,  $P_2 = 34\%$ ) y deseamos que el límite inferior del IC del resultado esté por encima de la mínima diferencia clínicamente relevante (9,5%), la diferencia entre ambos valores es del 5,5%, por lo que la amplitud del IC será del 11% ( $W_{IC} = 2 \times 0,055 = 0,11$ ). Aplicando la fórmula anterior, se obtiene que se requieren 748 pacientes por grupo.

Debido a que  $W_{IC}$  será habitualmente un valor pequeño, el tamaño de la muestra será superior al calculado con el método tradicional, el resultado será más preciso y con la suficiente potencia estadística. Además, este abordaje tiene la ventaja de facilitar la comunicación entre el clínico y el estadístico. Para el clínico es más difícil entender conceptos como el de la potencia estadística y el contraste de hipótesis que la interpretación de un IC.

Tal como hemos comentado en el apartado anterior, este enfoque es especialmente adecuado en el caso de los estudios de equivalencia y no-inferioridad.

**Ejemplo 15.11.** Retomemos el ejemplo 15.8 sobre un estudio hipotético que se propone evaluar la eficacia de una pauta monodosis para el tratamiento de las infecciones urinarias frente a la pauta habitual de 7 días, cuya eficacia se sitúa alrededor del 90% ( $P = 0,90$ ). Los investigadores habían establecido el margen de equivalencia (o de no-inferioridad) en el 10%. Dado que esperaban encontrar una diferencia del 5%, su IC debería tener una amplitud máxima que permitiera descartar el valor de dicho margen (un 5% por encima y por debajo del valor observado, por lo que  $W_{IC} = 2 \times 0,05 = 0,10$ ). Aplicando la fórmula, se determina que son necesarios 336 sujetos por grupo, estimación inferior a los 565 calculados por el procedimiento utilizado en el ejemplo 15.8.

De hecho, si se hubieran estudiado 336 sujetos por grupo y se hubiera obtenido una diferencia del 5% entre los grupos, los límites del IC 95% de esta diferencia serían 0,00013 y 0,09987 (es decir, 0,01% y 9,99%), lo que permite concluir la no-inferioridad de la monodosis (en caso de obtenerse estos resultados) ya que todo el IC queda por debajo del margen del 10% fijado.

Si se hubieran estudiado 565 sujetos por grupo, como se había calculado anteriormente, y se hubiera encontrado una diferencia del 5%, los límites del IC 95% en este caso serían del 1,2% y 8,8%. Este resultado permitiría llegar a la misma conclusión, pero habiendo estudiado un mayor número de sujetos.

## CORRECCIÓN POR LAS NO RESPUESTAS, LAS PÉRDIDAS Y LOS ABANDONOS

El número de sujetos calculado debe ampliarse en función del porcentaje de no respuestas, pérdidas y abandonos durante el seguimiento y de las desviaciones del protocolo que puedan esperarse, de forma que se asegure que finalizará el estudio el número de pacientes requerido. Una fórmula para hacerlo es la siguiente:

$$N_a = N [1/(1-R)]$$

donde  $N$  representa el número de sujetos teórico;  $N_a$ , el número de sujetos ajustado, y  $R$ , la proporción esperada de pérdidas.

**Ejemplo 15.12.** Para realizar un estudio se ha calculado que son necesarios 300 sujetos ( $N=300$ ) y se espera un 20% de pérdidas ( $R=0,20$ ). El número de pacientes que deberían iniciar el estudio sería  $N_a = 300 [1/(1 - 0,2)] = 375$  sujetos.

La utilización de esta fórmula asegura que el estudio mantenga la potencia estadística deseada, pero no evita que se puedan producir sesgos si las no respuestas, las pérdidas o los abandonos no se han producido aleatoriamente, es decir, si los sujetos que finalizan el estudio no son representativos de los que lo inician (lo que suele ser habitual).

## ESTRATEGIAS PARA MINIMIZAR EL NÚMERO DE SUJETOS NECESARIO

Estas estrategias se basan en conseguir una población homogénea, disminuir la variabilidad de las medidas y aumentar la frecuencia de aparición del

fenómeno de interés, por lo que deben aplicarse siempre que sea posible para aumentar la potencia del estudio (cuadro 15.5).

Algunas de estas estrategias enumeradas pueden implicar una modificación de las hipótesis de trabajo. Por ello, es importante realizar, ya en fases tempranas del diseño, una estimación aproximada del número de sujetos necesario, con el fin de evaluar la factibilidad del estudio y discutir si es necesario replantear o adaptar la hipótesis de trabajo o el estudio en general.

Cuadro 15.5 **Estrategias para disminuir el número de sujetos necesarios en un estudio**

- 1 Al seleccionar la variable de respuesta, usar:
  - Variables cuantitativas que aumentan la potencia de las pruebas estadísticas
  - Variables precisas
  - Variables de presentación frecuente
- 2 Al medir las variables, usar:
  - Promedios de medidas repetidas
  - Técnicas de medición precisas
  - Criterios y técnicas estandarizados
- 3 Seleccionar la población de estudio:
  - Homogéneas (criterios de selección restrictivos)
  - Poblaciones de alto riesgo, en las que es más frecuente observar la respuesta
- 4 Al diseñar el estudio, valorar si es posible:
  - Asignar aleatoriamente los sujetos a los grupos
  - Usar como medida de la respuesta las diferencias individuales entre el inicio y el final del estudio
  - Usar diseños especiales (p. ej., diseños cruzados)
  - Estratificar por factores pronósticos

## BIBLIOGRAFÍA DE LOS EJEMPLOS

Fayers PM, Cuschieri A, Fielding J, Craven J, Uscinska B, Freedman

LS. Sample size calculation for clinical trials: the impact of

clinician beliefs. Br J Cancer. 2000;82(1):213-9.

## BIBLIOGRAFÍA

- Bland JM. The tyranny of power: is there a better way to calculate sample size? *BMJ*. 2009;339:b3985.
- Campbell MJ, Julious SA, Altman DG. Estimating sample size for binary, ordered categorial, and continuous outcomes in two group comparison. *BMJ*. 1995;311:1145-8.
- Fitzner K, Heckinger E. Sample size calculation and power analysis: a quick review. *Diabetes Educ*. 2010;36(5):701-7.
- Florey CV. Sample size for beginners. *BMJ*. 1993;306: 1181-4.
- Kerry SM, Bland JM. Sample size in cluster randomisation. *BMJ*. 1998;316:549.
- Kosinski AS, Chen Y, Lyles RH. Sample size calculations for evaluating a diagnostic test when the gold standard is missing at random. *Stat Med*. 2010;29(15):1572-9.
- Hauschke D, Kieser M, Diletti E, Burke M. Sample size determination for proving equivalence based on the ratio of two means for normally distributed data. *Stat Med*. 1999;18:93-105.
- Marrugat J, Vila J, Pavesi M, Sanz F. *Estimación del tamaño de la muestra en la investigación clínica y epidemiológica*. Med Clin (Barc). 1998;111:267-76.
- Norman GR, Streiner DL. *Bioestadística*. Madrid: Mosby-Doyma; 1996.
- Obuchowski NA. Sample size calculations in studies of test accuracy. *Stat Methods Med Res*. 1998;7:371-92.
- Sahai H, Khurshid A. Formulae and tables for the determination of sample sizes and power in clinical trials for testing differences in proportions for the two-sample design: a review. *Stat Med*. 1996;15:1-21.
- Schlesselman JJ. Case-control studies: design, conduct, analysis. Nueva York: Oxford University Press; 1982.
- Tamayo-Sarver JH, Albert JM, Tamayo-Sarver M, Cydulka RK. Advanced statistics: how to determine whether your intervention is different, at least as effective as, or equivalent: a basic introduction. *Acad Emerg Med*. 2005;12:536-42.
- Whitehead J. Sample size calculations for ordered categorical data. *Stat Med*. 1993;12:2257-71.
- Wittes J. Sample size calculations for randomized controlled trials. *Epidemiol Rev*. 2002;24:39-53.

## Muestreo

El principio de representatividad implica que, para que los resultados de una investigación tengan validez interna, la muestra de sujetos estudiados debe ser representativa de la población de estudio. En los estudios transversales que estiman un parámetro se emplean técnicas de muestreo probabilístico, que aumentan la probabilidad de obtener una muestra representativa. Cuando no se dispone de un listado con los miembros que forman la población de estudio o cuando ésta es dinámica, y la muestra de sujetos se selecciona por métodos como la inclusión consecutiva de los pacientes que acuden a la consulta y cumplen los criterios de selección, para poder realizar inferencias, también debe asumirse que la muestra seleccionada es representativa de la población de estudio.

### VENTAJAS DEL MUESTREO

La razón principal para estudiar una muestra y no toda la población es el ahorro en tiempo, dinero y otros recursos. Es una cuestión de eficiencia. Si se ha calculado que son necesarios 300 individuos, y la población de referencia la forman 3.000 personas, estudiarlas a todas supone desperdiciar una buena cantidad de recursos, ya que, si la selección de la muestra y la medición de las variables se realizan correctamente, se obtendrá un resultado válido estudiando una muestra representativa de 300 personas.

Una segunda ventaja del muestreo, también relacionada con la validez del estudio, es que las mediciones tenderán a ser de mayor calidad, ya que se pueden dedicar más esfuerzos al estudiar un menor número de individuos. Paradójicamente, podría darse el caso de que se obtuvieran resultados

más fidedignos al estudiar una muestra que a toda la población.

También es un problema de cantidad de información. Al poder invertir más recursos en el estudio de cada individuo, pueden medirse otras variables, lo que no es posible hacer en toda la población. Si algunas de estas variables son potenciales factores de confusión, su medición posibilita su control en el análisis, lo que aumenta la validez de los resultados del estudio.

**Ejemplo 16.1.** En un estudio cuyo objetivo era conocer la prevalencia de factores de riesgo cardiovascular en jóvenes de 16 a 19 años, se seleccionó aleatoriamente una muestra de 1.176 estudiantes en los que se evaluaron unos pocos factores de riesgo (Andersen et al., 1989). A partir de esta muestra se seleccionó uno de cada cuatro estudiantes, en los que además se analizó su perfil lipídico, se midieron la presión arterial, el pliegue corporal, y el pulso en condiciones basales y después de una actividad intensa. Con la finalidad de establecer la representatividad de la segunda muestra se compararon la edad, el sexo, la talla y el peso con la primera, y no se encontraron diferencias significativas. El hecho de estudiar un conjunto de variables de forma más exhaustiva en una muestra reducida persigue el objetivo de alcanzar mayor exactitud y estandarización en las medidas llevadas a cabo.

### PLANIFICACIÓN DEL MUESTREO

Las *unidades de muestreo* son los elementos sobre los que se aplicará la técnica de selección. Pueden ser personas, centros de salud o escuelas, por ejemplo.

Si se quiere estimar la prevalencia de hipertensión arterial (HTA) en una población, serán las personas; si se pretende conocer el número de escuelas de una provincia en las que se imparte educación sanitaria, serán las escuelas.

Las unidades de muestreo no tienen que coincidir necesariamente con las *unidades de análisis*. En un estudio para conocer la prevalencia de HTA en la población escolar, las unidades de muestreo pueden ser las escuelas, pero se analiza en cada una la presión arterial de sus alumnos (unidades de análisis).

El primer paso para planificar un muestreo suele ser la definición del *marco muestral*, que es esencialmente una lista, o forma de identificación, de las unidades de muestreo sobre las que se aplicará el proceso de selección. La definición de este marco debe ser cuidadosa, ya que condiciona la interpretación de los resultados.

**Ejemplo 16.2.** En un estudio sobre la asociación entre la adaptación social y la depresión, Luty et al (2002) comprobaron que esta asociación no variaba entre los diferentes países, pero sí dependía del marco muestral del estudio, ya que se observaron resultados diferentes en la comunidad y en los centros sanitarios.

**Ejemplo 16.3.** Fuchs et al (2001) compararon la prevalencia de HTA que se obtenía al utilizar dos marcos muestrales diferentes: los adultos residentes en una población y los que acuden a una unidad de prevención de la HTA en un centro sanitario. Las diferencias en la prevalencia observada (el 24 frente al 42%, respectivamente) ilustran la influencia del proceso de muestreo sobre la evaluación de los factores de riesgo en la población.

El siguiente paso es decidir cómo seleccionar la muestra. En líneas generales, las técnicas de muestreo pueden clasificarse en probabilísticas y no probabilísticas.

## TÉCNICAS DE MUESTREO

### Muestreo probabilístico

Se define como el proceso de selección en que todos los individuos candidatos tienen una probabilidad conocida, distinta de cero, de ser incluidos en la muestra. Se utiliza alguna forma de selección aleatoria para obtener las unidades que serán estudiadas.

Existe una tendencia a considerar sinónimas las expresiones *muestra probabilística* y *muestra representativa*. El uso de una técnica de muestreo probabilístico

tiende a asegurar que se obtendrá una muestra representativa, en especial si la población y la muestra son de gran tamaño. Sin embargo, puede que no sea así, ya que el propio azar puede conducir a una muestra que no tenga la misma distribución de las variables de interés que la población de referencia, sobre todo si su tamaño es reducido. Por otro lado, pueden obtenerse muestras representativas utilizando técnicas no probabilísticas.

En todas las técnicas probabilísticas la selección de las unidades se realiza al azar y se evita la posible parcialidad, consciente o inconsciente, de los investigadores. Por esta razón, es más probable que las muestras tiendan a ser representativas, aunque ello no significa que estén exentas de error. En realidad, la muestra seleccionada no es más que una de las muchas que podrían obtenerse de la población de referencia. Es probable que la estimación de la variable obtenida a partir de una muestra difiera de su verdadero valor en la población de referencia por azar (error aleatorio o debido al muestreo), y lo es más cuanto menor es el tamaño de la muestra. El uso de muestras probabilísticas permite la aplicación de técnicas estadísticas capaces de cuantificar este error aleatorio.

### Muestreo aleatorio simple

Es aquella técnica en la que cada unidad del marco muestral tiene la misma probabilidad de ser escogida y en la que cada una de las posibles muestras del mismo tamaño tienen la misma probabilidad de ser escogidas. El muestreo aleatorio simple es, pues, un procedimiento equiprobabilístico, pero no todo método equiprobabilístico es aleatorio simple.

**Ejemplo 16.4.** Silva (1993) ilustra este concepto con el siguiente ejemplo: supongamos un grupo de 10 niños ordenados alfabéticamente, de los que se quiere seleccionar una muestra de 5 niños. Consideremos los tres siguientes procedimientos de selección: *a*) tomar los 5 primeros de la lista; *b*) dividir el listado en dos grupos (los 5 primeros y los 5 últimos), lanzar una moneda al aire y tomar como muestra los elementos del primer grupo en caso de obtener cara, y los del segundo en caso de obtener cruz, y *c*) numerar los 10 elementos y obtener 5 números aleatorios. El método *a* no es probabilístico, ya que los últimos 5 niños del listado no podrán salir en la muestra. El método *b* es equiprobabilístico porque todos los niños tienen la misma probabilidad de salir en la muestra, pero no es aleatorio simple, ya que sólo dos muestras pueden ser elegidas (existen subconjuntos de 5 elementos que no podrían ser seleccionados). El método *c* es aleatorio simple, ya

que cualquiera de los 252 subconjuntos diferentes de 5 niños puede ser seleccionado y con la misma probabilidad.

Como se deduce del ejemplo anterior, el primer paso es preparar una lista de las unidades de muestreo, numerándolas, por ejemplo, de forma secuencial. Este proceso puede ser fácil y rápido si la población de referencia se encuentra en una base de datos informatizada, o muy laborioso si se realiza manualmente y la población es amplia. A continuación, se seleccionan tantos números aleatorios como elementos debe tener la muestra. La selección de números aleatorios suele realizarse con un ordenador o una calculadora de mano.

Esta técnica es sencilla y de fácil comprensión, y permite el cálculo rápido de medias y variancias. Requiere, sin embargo, que se conozca de antemano la lista completa de todas las unidades de muestreo. Si la lista contiene fallos u omite segmentos de la población, se perderán las ventajas propias del muestreo aleatorio, ya que todas las unidades no tendrán la misma probabilidad de aparecer en la muestra.

### Muestreo aleatorio estratificado

Se trata de una modificación del método anterior que intenta asegurar que la muestra presente la misma distribución que la población en relación con determinadas variables.

La población se divide en estratos en función de las categorías de las variables por las que se desea estratificar, formando subgrupos con alguna característica en común y mutuamente excluyentes. A continuación, se escoge una muestra aleatoria de cada estrato, manteniendo las proporciones observadas en la población de referencia (*muestreo aleatorio estratificado proporcional*).

Los estratos deben definirse en función de variables que puedan influir sobre los resultados. Si se desea determinar la prevalencia de HTA en una comunidad, posibles factores de confusión serían, entre otros, la edad, el sexo o la presencia de obesidad, ya que son variables relacionadas con la presencia de HTA. Si se aplicara un muestreo aleatorio simple, podría obtenerse una muestra en la que la distribución de estas variables no fuera la misma que en la población, lo que conduciría a un sesgo en los resultados. Un muestreo estratificado por estos factores garantizaría una distribución similar a la de la población de referencia.

**Ejemplo 16.5.** Supongamos un marco de muestreo de 2.000 personas en las que se desea estimar la prevalencia de HTA, estudiando para ello una muestra de 250 sujetos. Se conoce que la prevalencia de HTA varía en función del sexo y la presencia de obesidad. Por tanto, los investigadores desean que la muestra en que se realizará el estudio refleje fielmente la distribución, según el sexo y la presencia de obesidad que tiene la población de referencia, que se muestra en la [tabla 16.1](#). Así, por ejemplo, existe un 12% de hombres obesos, por lo que, de los 250 sujetos de la muestra, un 12% deberá tener estas características. Por tanto, de los 240 hombres obesos de la población, deben seleccionarse aleatoriamente 30 para formar parte de la muestra. Se procede de la misma forma con los otros estratos ([tabla 16.1](#)).

La correcta aplicación de esta técnica requiere que se conozca la distribución en la población de la(s) variable(s) empleada(s) para la estratificación y que la variabilidad entre los distintos estratos sea lo más amplia posible, mientras que dentro de cada estrato debe ser lo más pequeña posible. Así, se obtendrán subgrupos homogéneos respecto a la posible variable

**Tabla 16.1 Muestreo estratificado (ejemplo 16.6)**

Distribución de una población de 2.000 personas en función del sexo y la obesidad

Varones	Mujeres	Total
Obesidad	240 (12%)	560 (28%)
No obesidad	960 (48%)	1.440 (72%)
Distribución de una muestra de 250 personas que mantiene la misma distribución en función del sexo y la obesidad que la población general		
Varones	Mujeres	Total
Obesidad	40 (16%)	70 (28%)
No obesidad	120 (48%)	180 (72%)

de confusión (todos los miembros de un estrato la presentan o no) y heterogéneos entre sí (miembros de distintos estratos no comparten la variable).

En general, con el muestreo estratificado se obtienen estimaciones más precisas que con el muestreo aleatorio simple, porque la variancia total se basa en la de cada uno de los estratos, y éstos son más homogéneos que la población. El aumento de precisión que se consigue depende de las diferencias entre los estratos. Cuanto mayor sea esta diferencia, y menor la variabilidad dentro de cada estrato, mayor será la reducción de la variancia que se produce. Por tanto, siempre que se utiliza un muestreo estratificado, debe tenerse en cuenta en el análisis, al realizar las estimaciones.

Cuando alguno de los estratos tiene un tamaño reducido, puede desaparecer esta ventaja, ya que disminuiría la precisión de la estimación en dicho estrato y podría afectar a la de la estimación global. En esta situación, puede aplicarse un *muestreo no proporcional* que sobrerepresente este estrato, aplicando sobre él una fracción de muestreo diferente. Esta alternativa dificulta el análisis, ya que la estimación del parámetro en el global de la muestra deberá hacerse a partir de cada estrato por separado, y ponderando la diferente variabilidad de cada uno de ellos. De hecho, el muestreo no proporcional se utiliza más cuando la finalidad es realizar comparaciones entre los estratos, con el fin de asegurar que cada uno de ellos tiene el tamaño suficiente para obtener estimaciones con la suficiente precisión.

**Ejemplo 16.6.** Supongamos que se lleva a cabo una encuesta entre los médicos de atención primaria para conocer su opinión sobre un determinado problema de salud. La distribución de médicos según su ámbito de trabajo es la siguiente: 230 (56%) trabajan en el medio urbano, 160 (39%) en el semiurbano y 20 (5%) en el rural. Si se seleccionara una muestra de 100 médicos y se mantuvieran las mismas proporciones, sólo se conseguirían las respuestas de 5 médicos rurales, lo cual sería insuficiente para conocer su opinión y compararla con la de otros grupos. Los investigadores deciden aplicar diferentes fracciones de muestreo en cada uno de los estratos (urbanos, 20%; semiurbanos, 30%; rurales, 100%), con lo que la distribución de encuestados es de 46, 48 y 20, respectivamente. Hay que tener en cuenta las diferentes fracciones si se quieren realizar estimaciones globales de toda la muestra, lo que añade de cierta dificultad al análisis. Si se tiene la intención de realizar un análisis sólo de los subgrupos por separado, no es necesario tener en cuenta las diferentes fracciones de muestreo, porque cada estrato puede considerarse una población diferente.

## Muestreo en múltiples etapas

Esta técnica consiste en seleccionar unidades de muestreo de una población (unidades primarias) y, en una segunda etapa, obtener una muestra de cada una de las unidades primarias (unidades secundarias). Por ejemplo, si se desea estudiar la prevalencia de escoliosis en niños de edad escolar, una manera de seleccionar la muestra sería elegir al azar escuelas y, dentro de cada una, una muestra de niños. Se pueden usar el número de etapas que sean necesarias y, en cada una de ellas, aplicar un método diferente de muestreo (simple, estratificado, sistemático). Cuando en el estudio se incluyen todas las unidades secundarias (es decir, la fracción de muestreo es del 100%), se denomina *muestreo en conglomerados*.

Esta técnica es útil cuando la población de referencia es muy grande y está dispersa, como en el caso de la población escolar de un país o una gran ciudad. El uso de una técnica aleatoria simple sería muy costoso, ya que la muestra incluiría a pocos estudiantes de cada institución y los investigadores tendrían que ir a casi todas ellas para entrevistarlos o examinarlos, con lo que se rentabilizaría muy poco cada viaje. Además, es probable que no se disponga de un listado de todos los niños en edad escolar de una ciudad, pero sí de una lista de escuelas, y que cada una de ellas tenga el listado de sus alumnos.

Su desventaja principal es que si las unidades primarias contienen personas similares en relación con el fenómeno de interés, la precisión de la estimación será menor que si se utiliza un muestreo aleatorio simple, por ejemplo. Si se considera un estudio sobre los hábitos alimentarios de una comunidad y se encuesta a 100 individuos obtenidos mediante un muestreo aleatorio simple, el grado de información y la precisión de las estimaciones no será igual que si se estudian 100 individuos (unidad secundaria) procedentes de 25 familias (unidad primaria). Los hábitos alimentarios son muy parecidos entre miembros de una misma familia, por lo que, en realidad, es casi como si sólo se estudiara a 25 personas. Además, si una o varias familias son muy numerosas, contribuirán de una forma desproporcionada al tamaño de la muestra.

Otro inconveniente es que la variabilidad es mayor que las técnicas anteriores, y su cálculo es complejo. Una solución parcial es extraer una muestra de mayor tamaño y considerarla como aleatoria simple, y procurar que las unidades de muestreo sean lo más pequeñas posible.

**Ejemplo 16.7.** Un estudio tenía por objetivo conocer las creencias y opiniones de los escolares respecto al tabaco, el alcohol y el ejercicio físico (Comín et al, 1989).

Se seleccionó una muestra representativa de escolares de la ciudad de Barcelona. La técnica de muestreo empleada fue la de conglomerados en varias etapas. Se tomó como unidad de muestreo el aula, estratificando por las siguientes variables: curso, tipo de escuela (pública, privada) y tamaño del curso (número de aulas). Es importante remarcar que todos estos factores están relacionados con las variables estudiadas; el curso se asocia a los hábitos de vida a través de la edad, y el tipo de escuela, a través de la clase social. La estratificación permite obtener estimaciones más precisas. El empleo del muestreo por conglomerados es prácticamente inevitable al estudiar la población escolar. El aumento de la variabilidad que acompaña a la utilización de conglomerados se minimiza al tomar el aula, en lugar del colegio, como unidad primaria de muestreo, y se compensa parcialmente con la estratificación.

## Muestreo sistemático

Este procedimiento se basa en aplicar alguna regla sistemática simple, como elegir uno de cada  $n$  individuos. Los pasos a seguir son los siguientes: en primer lugar, se calcula la *constante de muestreo* ( $k$ ), dividiendo el total de la población candidata por el tamaño de la muestra deseado. A continuación, se extrae la primera unidad al azar entre las  $k$  primeras unidades de muestreo, y se le suma la constante, sucesivamente, hasta completar el tamaño de la muestra.

**Ejemplo 16.8.** Supongamos que deseamos obtener una muestra de 200 historias clínicas de un archivo de 1.000 historias. La constante de muestreo es  $1.000/200 = 5$ , lo que significa que se escoge una de cada 5 historias. La primera se elige de forma aleatoria entre los números del 1 al 5. Si la elegida es la 2, la siguiente es la 7 ( $2+5$ ), después la 12 ( $7+5$ ), la 17 ( $12+5$ ), y así sucesivamente, hasta conseguir las 200 historias necesarias.

La primera ventaja del muestreo sistemático es que es fácil de aplicar y, en ocasiones, es más cómodo y práctico que el muestreo aleatorio simple. Es más sencillo, por ejemplo, escoger todas las historias cuyo número termina en cero que a partir de un listado de números aleatorios. La segunda es que no siempre es necesario tener de antemano una lista completa y exhaustiva de toda la población. Los sujetos se pueden ir seleccionando a medida que el estudio avanza, siempre y cuando lleguen en algún orden preestablecido, como por ejemplo, los pacientes que acuden a una consulta. La tercera ventaja es que, cuando la población de referencia

está ordenada siguiendo una tendencia conocida (de mayor a menor, de más viejo a más joven, etc.), el muestreo sistemático asegura una cobertura de unidades de todos los tipos. Si interesa extraer una muestra de los médicos de una provincia y el colegio profesional tiene las listas ordenadas por el año de licenciatura, podría darse el caso de que con un muestreo aleatorio simple predominaran los médicos de una determinada edad. Sin embargo, con un muestreo sistemático, se aseguraría una muestra que representaría todos los años de licenciatura.

La limitación más importante para su uso aparece cuando la constante que se aplica está relacionada con el fenómeno que se quiere medir. En este caso se pueden hallar estimaciones muy distorsionadas de los parámetros poblacionales.

**Ejemplo 16.9.** Supongamos que se quiere estudiar la actividad asistencial de un equipo de atención primaria y que la unidad de muestreo son los días de la semana. Si el primer día seleccionado fuera un lunes y se aplicara una constante ( $k=7$ ), se recogería únicamente la actividad realizada los lunes de cada semana. Con este procedimiento se sobreestimaría la media real de visitas, ya que tradicionalmente los lunes son los días de mayor actividad. Como puede deducirse, estas situaciones son muy poco frecuentes y, en todo caso, es relativamente fácil intuir que la constante de muestreo está relacionada con el fenómeno de estudio.

Las ventajas y desventajas de cada uno de los métodos de muestreo probabilístico se encuentran en la tabla 16.2.

## Muestreo no probabilístico

En el muestreo no probabilístico se desconoce la probabilidad que posee cada unidad de ser incluida en la muestra, y la selección se realiza mediante métodos en los que no interviene el azar.

El uso de pruebas estadísticas requiere que la muestra analizada sea aleatoria. Su aplicación en muestras no obtenidas por un método aleatorio se hace asumiendo que están libres de sesgos y que son representativas de la población de referencia.

La técnica más frecuente es el *muestreo consecutivo*. Consiste en seleccionar a los pacientes que cumplen los criterios de selección especificados en el protocolo del estudio, a medida que acuden a la consulta durante un período determinado.

**Ejemplo 16.10.** Lee et al (2002) compararon las características de 1.546 pacientes que acudieron consecutivamente a un centro de salud con las de

**Tabla 16.2 Ventajas y desventajas de las principales técnicas de muestreo probabilístico**

Ventajas	Desventajas
<b>Muestreo aleatorio simple</b>	
<ul style="list-style-type: none"> <li>- Sencillo y de fácil comprensión</li> <li>- Cálculo rápido de medias y variancias</li> </ul>	<ul style="list-style-type: none"> <li>- Requiere que se posea de antemano un listado completo de toda la población</li> <li>- Cuando se trabaja con una muestra pequeña, es posible que no sea representativa de la población</li> </ul>
<b>Muestreo aleatorio estratificado</b>	
<ul style="list-style-type: none"> <li>- Tiende a asegurar que la muestra representa adecuadamente a la población en función de unas variables determinadas</li> <li>- Se obtienen estimaciones más precisas</li> <li>- Se pueden aplicar distintas fracciones de muestreo en cada estrato</li> </ul>	<ul style="list-style-type: none"> <li>- Se ha de conocer la distribución en la población</li> <li>- Requiere que se posea de antemano un listado completo de toda la población</li> <li>- Cálculo de las estimaciones más complejo que en el muestreo aleatorio simple</li> </ul>
<b>Muestreo en etapas múltiples</b>	
<ul style="list-style-type: none"> <li>- Es muy eficiente cuando la población es muy grande y dispersa</li> <li>- La necesidad de listas de las unidades de una etapa se limita a aquellas unidades de muestreo seleccionadas en la etapa anterior</li> </ul>	<ul style="list-style-type: none"> <li>- Las estimaciones son menos precisas que con el muestreo aleatorio simple o estratificado</li> <li>- El cálculo de las estimaciones es complejo</li> </ul>
<b>Muestreo sistemático</b>	
<ul style="list-style-type: none"> <li>- Fácil de aplicar</li> <li>- No siempre es necesario disponer de una lista de toda la población</li> <li>- Cuando la población está ordenada siguiendo una tendencia conocida, asegura una cobertura de unidades de todos los tipos</li> </ul>	<ul style="list-style-type: none"> <li>- Si la constante de muestreo está asociada con el fenómeno de interés, se pueden obtener estimaciones sesgadas</li> </ul>

la población de referencia. Observaron que los pacientes incluidos en la muestra consultaban con mayor frecuencia, eran mayores, con peor estado de salud (mayor prevalencia de HTA, accidente cerebrovascular y cáncer), menos fumadores y vivían solos, con mayor frecuencia.

El muestreo consecutivo es la técnica más utilizada en los ensayos clínicos. Si se evalúa la eficacia de una medida terapéutica contra la gastroenteritis, no se dispone al inicio del estudio de un listado de la población que desarrollará la enfermedad, por lo que no puede aplicarse un muestreo aleatorio simple. En esta situación, el equipo investigador se limita a reclutar a los individuos que cumplen los criterios de selección a medida que son diagnosticados.

La inclusión de *voluntarios* es otra técnica de muestreo no probabilístico. En general, es preferible rechazar su colaboración, ya que las verdaderas razones que llevan a una persona a prestarse para un estudio sin haber sido seleccionado para ello suelen estar relacionadas con otras características que afectan a la representatividad de la muestra.

Otra técnica utilizada en encuestas de opinión y de mercado, pero muy poco en la investigación epidemiológica, es el *muestreo por cuotas*, en el que la composición general de la población de referencia en términos de características como la edad, el sexo, o la clase de educación se decide, o se conoce, de antemano. Seguidamente, se determina el número de personas requerido, o las cuotas, para los hombres y las mujeres, según las diferentes edades y la clase de educación, y el único requerimiento es conseguir el número adecuado de personas para llenar cada una de las cuotas.

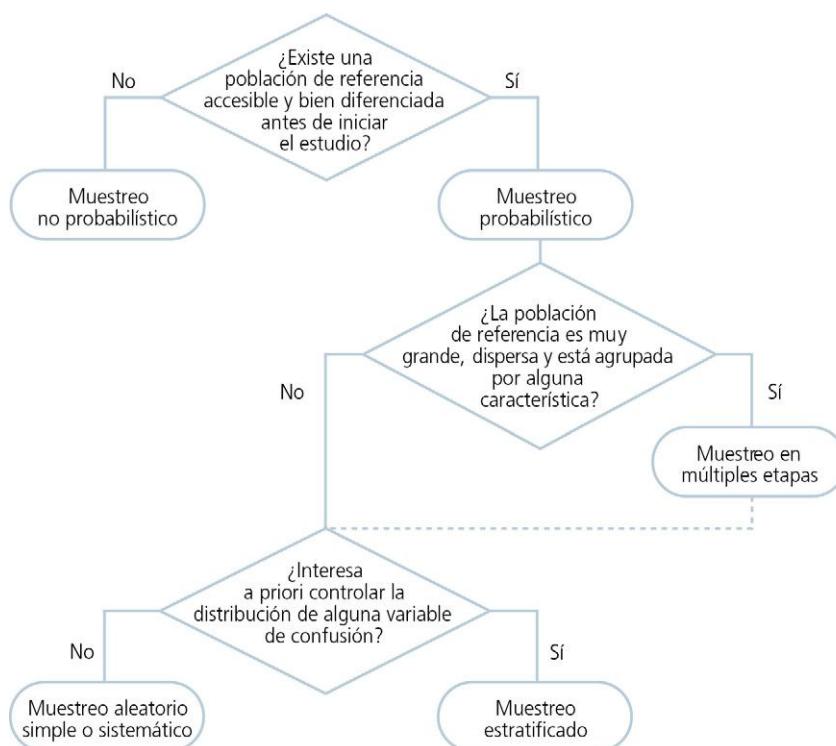
En algunas ocasiones, en especial cuando tanto el tamaño de la población de referencia como el de la muestra son muy pequeños, la selección puede hacerse por criterios razonados. Por ejemplo, si se desea conocer el funcionamiento de los centros de salud de una provincia a través de un análisis exhaustivo de dos de ellos, es preferible seleccionar los dos centros que puedan considerarse los más representativos, en vez de dejar la selección al azar.

En las *técnicas adaptativas* el diseño del muestreo se adapta en función de las observaciones que se

van realizando durante el estudio, por ejemplo, solicitando a los participantes en un estudio sobre consumo de drogas que informen al investigador sobre otros consumidores. Dependiendo del mecanismo de adaptación, reciben nombres como *bola de nieve*, *paseos aleatorios* o *muestreo en red*. Estas técnicas son especialmente útiles cuando la población de interés es difícil de alcanzar, ya que, además de la

ganancia que suponen en términos de eficiencia en relación con las técnicas convencionales, pueden ser la única forma operativa de obtener una muestra lo suficientemente amplia. Su principal inconveniente es que deben aplicarse procedimientos estadísticos que tengan en cuenta su utilización.

La figura 16.1 presenta un algoritmo para la selección de la técnica de muestreo.



Las preguntas son orientativas y han de realizarse en función de las ventajas y desventajas de cada técnica

**Figura 16.1** Algoritmo para la selección de la técnica de muestreo.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

Andersen LB, Henckel P, Saltin B. Risk factors for cardiovascular disease in 16-19 year old teenagers. *Int J Med.* 1989;225:157-63.

Comín E, Nebot M, Villalbí JR. Exercici i consum de tabac i alcohol dels escolars de Barcelona. *Gac Sanit.* 1989;3:355-65.

Fuchs SC, Petter JG, Accordi MG, Zen VL, Pizzol Jr AD, Moreira

LB, et al. Establishing the prevalence of hypertension. Influence of sampling criteria. *Arq Bras Cardiol.* 2001;76: 445-52.

Lee ML, Yano EM, Wang M, Simon BF, Rubenstein LV. What patient population does visit-based sampling in primary care settings represent? *Med Care.* 2002;40:761-70.

Luty SE, Joyce RR, Mulder RT, McKenzie JM. Social adjustment in depression: the impact of depression severity, personality, and clinic versus community sampling. *J Affect Disord.* 2002;70:143-54.

Silva Ayçaguer LC. Muestreo para la investigación en ciencias de la salud. Madrid: Díaz de Santos; 1993.

## BIBLIOGRAFÍA

- Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. 4th ed. Oxford: Blackwell Science; 2002.
- Kelsey JL, Wittemore AS, Evans A, Thompson WD. Methods in observational epidemiology.
- 2nd ed. New York: Oxford University Press; 1996.
- Phillips CV, LaPole LM. Quantifying errors without random sampling. BMC Med Res Methodol. 2003;3:9.
- Rodríguez Osuna J, Métodos de muestreo. Colección «Cuadernos metodológicos» núm. 1. Madrid: Centro de Investigaciones Sociológicas; 1991.
- Thompson SK, Collins LM. Adaptive sampling in research on risk-related behaviors. Drug Alcohol Depend. 2002;68(Suppl 1):S57-67.

## Formación de los grupos de estudio

### PRINCIPIO DE COMPARABILIDAD

Los estudios analíticos tratan de estimar el efecto de un factor de estudio sobre una variable de respuesta, o la magnitud de la asociación entre ambas variables, mediante la comparación de los resultados observados en un grupo de estudio con los hallados en otro grupo que actúa de referencia o control. La función del grupo control es proporcionar una estimación del valor de la variable de respuesta en ausencia del factor de estudio. Para que esta estimación sea una referencia válida, el grupo control debe ser comparable al de estudio en todas aquellas variables que puedan influir sobre la respuesta, de forma que permita aislar el efecto del factor de estudio del causado por otros factores.

Por tanto, la formación de los grupos debe realizarse de forma que no difieran significativamente en la distribución de las variables pronósticas, excepto –obviamente– en la del factor de estudio. La comparabilidad de los grupos, desde el inicio del estudio hasta su final, es fundamental, ya que de ella dependerá, en gran medida, la validez de las conclusiones.

Para intentar conseguir la comparabilidad de los grupos deben buscarse tres objetivos: que los sujetos tanto del grupo de estudio como del grupo control sean representativos de una misma población de referencia; que las variables pronósticas se distribuyan de forma equilibrada en ambos grupos, y que se obtenga información válida de todos los sujetos de ambos grupos. Sus finalidades son, respectivamente, eliminar o reducir el sesgo de selección, controlar los factores de confusión y eliminar o reducir el sesgo de información.

Existen algunas técnicas que pueden ser útiles para intentar asegurar la comparabilidad inicial de los grupos. La *asignación aleatoria* de los sujetos a los grupos de estudio se basa en la tendencia que tiene el azar de formar grupos que presenten distribuciones similares de todos los factores conocidos o no. Esta técnica sólo es utilizable en los estudios experimentales. El *emparejamiento (matching)* consiste en seleccionar, para cada individuo del grupo de estudio, uno o varios sujetos de referencia similares en las características que se desean controlar, y se emplea preferentemente en los estudios de casos y controles. Estas estrategias suelen complementarse con la *restricción* de los criterios de selección, de forma que tanto el grupo de estudio como el de control sean homogéneos.

Teniendo en cuenta el principio de comparabilidad, lo más adecuado es que el grupo control sea *contemporáneo* o *concurrente*, es decir, que la medición de las variables se realice durante el mismo período de tiempo que en los sujetos del grupo de estudio. El uso de otros tipos de grupo control tiene ciertas limitaciones.

Por *controles históricos* se entiende un grupo de pacientes que se asume que tuvieron la misma enfermedad o condición que el grupo que se va a estudiar, pero que fueron diagnosticados, tratados o seguidos en un espacio de tiempo anterior. Su inconveniente es que es difícil asegurar que los criterios de selección y seguimiento de los sujetos, de medición y evaluación de la respuesta, y la distribución de las variables pronósticas son similares a las del grupo de estudio (cuadro 17.1).

Otros tipos de grupo control que pueden utilizarse en algunas ocasiones son las series de pacientes obtenidos de la literatura científica, de bases de datos o a partir de su participación en algún otro estudio,

### Cuadro 17.1 Argumentos en contra de la utilización de controles históricos

- Posibilidad de sesgos de selección
- Posibilidad de que se hayan presentado cambios en los criterios diagnósticos, exploraciones, etc., durante el tiempo transcurrido
- Dificultades para el cálculo de las pruebas de significación estadística
- Menor capacidad de convencimiento de los profesionales sobre la validez de las conclusiones
- Dificultades para obtener financiación

o incluso datos correspondientes a la población general. Estos grupos tienen limitaciones similares a las comentadas para los controles históricos.

## COMPARABILIDAD DE LOS GRUPOS EN ESTUDIOS OBSERVACIONALES

El proceso de formación de los grupos depende del tipo de estudio. En los diseños observacionales se realiza en función de la existencia o ausencia de la enfermedad de interés (en los estudios de casos y controles) o de la presencia o ausencia de la exposición (en los estudios de cohortes). En los estudios experimentales la situación es muy distinta, ya que los sujetos se asignan a los diferentes grupos que se desea comparar por un procedimiento aleatorio.

Dado que en los estudios observacionales el investigador no decide a qué grupos serán asignados los sujetos, pueden aparecer limitaciones importantes en la comparabilidad inicial.

### Estudios de cohortes

En los estudios en los que se define una cohorte general cuyos sujetos se clasifican en expuestos (cohorte de estudio) y no expuestos (cohorte interna de comparación), puede asegurarse que ambos grupos proceden de la misma población de estudio (cohorte general) (fig. 17.1). Esta situación es muy similar a la de los ensayos clínicos, excepto que, en los estudios de cohortes, los sujetos no son asignados a los grupos al azar, sino que se clasifican en función de su exposición al factor de estudio. Dado que esta exposición puede estar asociada con otros factores pronósticos (potenciales factores de confusión), puede afectarse la comparabilidad inicial de los grupos. Por lo tanto, deben medirse todas las variables pronósticas conocidas e intentar su influencia en la fase de análisis para obtener resultados válidos. En cambio, en los ensayos clínicos aleatorios, al existir una asignación al azar de los sujetos a los grupos, si el tamaño de la muestra es grande, se tiende a conseguir grupos comparables por todos los posibles factores de confusión, conocidos o no, medida o no. Además, la comparabilidad entre los grupos también puede afectarse por las diferencias en la recogida de la información y por las pérdidas de sujetos que pueda haber durante el seguimiento.

En los diseños en que no se parte de una cohorte general, sino que se identifica un grupo de sujetos

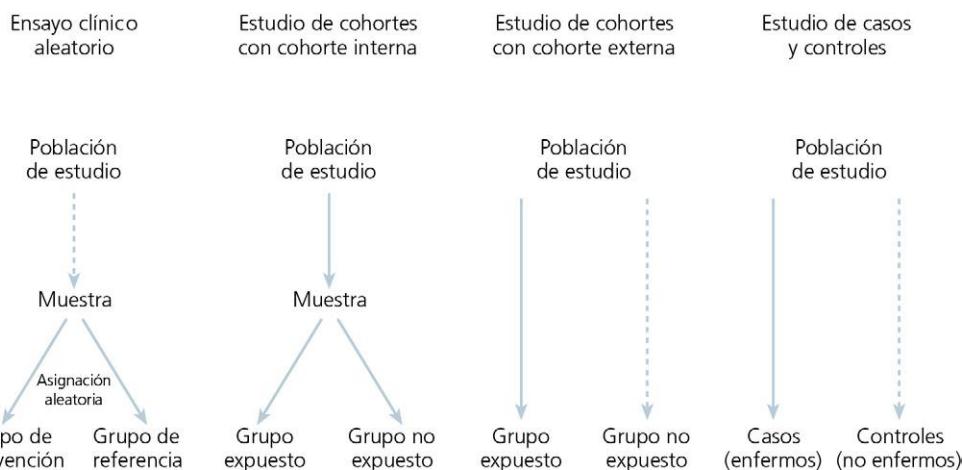


Figura 17.1 Formación de los grupos en los diseños analíticos.

expuestos al factor de estudio (cohorte de estudio) y un grupo de sujetos no expuestos que actúa como control (cohorte externa de comparación) (ver fig. 17.1), a los problemas comentados anteriormente se les añade la dificultad de asegurar que la cohorte externa sea representativa de la misma población de la que procede la cohorte de estudio, es decir, de que ambas procedan de una misma población.

En los estudios de cohortes también puede utilizarse el emparejamiento: para cada sujeto expuesto incluido se seleccionan uno o más individuos no expuestos que tengan el mismo valor para las variables cuyo potencial efecto de confusión se desea controlar.

**Ejemplo 17.1.** Ludvigsson et al (2007) estudiaron la relación entre la enfermedad celíaca y el riesgo de tuberculosis. Para ello, llevaron a cabo un estudio de cohortes de base poblacional, en que la cohorte expuesta estaba formada por los 14.336 pacientes celíacos entre 1964 y 2003. Para cada sujeto expuesto, se seleccionaron 5 controles emparejados por edad, sexo, año y área geográfica, de manera que estas variables quedaron distribuidas de forma equilibrada en ambas cohortes, previniendo así su posible efecto de confusión. Los autores encontraron un riesgo de tuberculosis aumentado entre los pacientes con enfermedad celíaca (riesgo relativo [RR] = 3,7; intervalo de confianza del 95% [IC 95%]: 2,1-6,5).

En los estudios de cohortes, el emparejamiento no requiere aplicar procedimientos estadísticos específicos, a diferencia de lo que ocurre en los estudios de casos y controles, a no ser que se produzcan pérdidas en el seguimiento que puedan introducir un desequilibrio entre los grupos.

## Estudios de casos y controles

El grupo de casos y el de controles deben ser representativos de una misma población, formada conceptualmente por los miembros de una cohorte subyacente, definida por los criterios de inclusión y exclusión (ver fig. 17.1). Un tipo de estudio paradigmático en este sentido es el de casos y controles anidado en una cohorte, donde se identifica claramente la cohorte subyacente, de la que proceden tanto los casos como los controles.

En los estudios de casos y controles poblacionales, en los que se seleccionan todos los casos que se desarrollan en una zona determinada, el grupo control debe ser una muestra representativa de la población que reside en el área geográfica de la que proceden los casos. En estas circunstancias, la selección del grupo control se simplifica, ya que puede extraerse

una muestra aleatoria de la población mediante técnicas probabilísticas.

En cambio, cuando el diseño no es poblacional, sino que los casos se detectan sin identificar la cohorte subyacente de donde proceden, el grupo control debe ser representativo de aquellos individuos que, en el supuesto de desarrollar la enfermedad, hayan sido detectados como casos. Si no se tienen en cuenta estos supuestos, se podría cometer un sesgo de selección relacionado con alguna variable ligada al hecho de poder ser detectado como caso. Así, en un estudio de casos y controles hospitalarios, la selección de los controles a partir de los individuos ingresados en el mismo hospital en que se han identificado los casos se sustenta en la asunción de que la distribución de la exposición es la misma que en la población de donde proceden los casos. Esta asunción es razonable cuando se cumplen dos supuestos:

- *El área de influencia del hospital es la misma para los casos y para los controles.* Implica que los sujetos que son ingresados en un hospital como casos habrían sido ingresados en el mismo hospital si hubieran tenido la enfermedad de los controles, y viceversa, aquellos ingresados como controles habrían sido ingresados en el mismo hospital en el supuesto de que hubiesen padecido la enfermedad de los casos.
- *La selección de los controles no está relacionada con la exposición en estudio.* En la práctica, implica excluir del grupo control a los pacientes que tengan enfermedades relacionadas de forma positiva o negativa con la exposición en estudio. Igualmente, con el fin de evitar que los controles puedan concentrarse en alguna enfermedad relacionada con la exposición, conviene escoger controles con distintas patologías, con el fin de minimizar un posible sesgo debido al muestreo.

Ante la disyuntiva de cuál es el mejor grupo control, y teniendo en cuenta que cada alternativa tiene ventajas e inconvenientes, en muchas ocasiones se decide incluir dos o más grupos control. Sin embargo, si se obtienen resultados inconsistentes entre los diferentes grupos, el análisis y la interpretación de los resultados pueden volverse muy complejos.

**Ejemplo 17.2.** Pomp et al (2010) utilizaron los datos del estudio MEGA (*Multiple Environmental and Genetic Assessment of risk factors for venous thrombosis*), un estudio de casos y controles de base poblacional de gran tamaño, para estudiar la idoneidad de diferentes grupos control. Los casos correspondían a los sujetos diagnosticados de un primer episodio

de trombosis venosa entre 1999 y 2004. Se incluyeron dos grupos control: uno formado por los cónyuges de los casos, y otro obtenido por un procedimiento de selección telefónica aleatoria. Ambos grupos control presentaron índices de masa corporal más bajos que los casos, aunque se aproximaron más en los cónyuges, probablemente relacionado con estilos de vida similares, lo que condujo a estimaciones de la *odds ratio* (OR) más cercanas a la unidad. En cambio, cuando se estudiaron las lesiones como factor de riesgo, la OR de los cónyuges era superior, incluso realizando un análisis apareado. Evaluaron también el factor V de Leiden como ejemplo de factor genético, y encontraron una frecuencia idéntica en ambos grupos, por lo que podían ser combinados en un único análisis. Los autores concluyen que, aunque las medidas del efecto con ambos grupos control iban en la misma dirección y con un orden de magnitud similar, no era siempre el mismo grupo el que producía las estimaciones más elevadas o reducidas, y que un análisis apareado no solventaba la situación.

## COMPARABILIDAD DE LOS GRUPOS EN LOS ESTUDIOS EXPERIMENTALES: ASIGNACIÓN ALEATORIA

En los estudios experimentales, el investigador debe distribuir a los sujetos en dos grupos que sean comparables por los factores pronósticos. La *asignación aleatoria*, también conocida por el anglicismo *randomización*, es la técnica que más tiende a asegurar esta comparabilidad.

La asignación aleatoria significa que la decisión de qué tratamiento (o intervención) recibirá cada uno de los pacientes incluidos en el estudio se realiza al azar, sin que exista ninguna influencia por parte del propio sujeto o del investigador. Por ello, el ensayo clínico aleatorio es el estudio que proporciona mayor fuerza a los resultados observados, y constituye el estándar con el que se compara el resto de los diseños de investigación.

### Ventajas de la asignación aleatoria

Las ventajas de la asignación aleatoria son múltiples (cuadro 17.2). En primer lugar, tiende a producir una distribución equilibrada de las variables pronósticas, tanto de las conocidas como de las desconocidas, entre los diferentes grupos de estudio. Si el proceso es realmente aleatorio, ni el sujeto ni el investigador

### Cuadro 17.2 Ventajas de la asignación aleatoria

- Tiende a producir una distribución equilibrada de las variables pronósticas, conocidas y no conocidas
- Previene sesgos debidos al propio sujeto y al observador
- Permite la utilización de las pruebas estadísticas
- Permite el uso de las técnicas de enmascaramiento

influyen en la decisión de qué intervención recibirá cada uno de los participantes. De esta forma se intentan prevenir posibles sesgos que puedan comprometer la comparabilidad de los grupos.

El azar *tiende* a producir grupos comparables, pero no lo asegura. Esta tendencia es mayor cuanto más elevado es el número de sujetos. Por consiguiente, en los estudios que se llevan a cabo con un número de pacientes relativamente reducido, aumenta la probabilidad de que alguna variable quede distribuida de forma desigual. Si esta variable está relacionada con la respuesta que se mide, actuará como un factor de confusión.

La asignación aleatoria permite el uso del cálculo de probabilidades para determinar hasta qué punto una diferencia observada puede ser causada por el azar, para garantizar que las pruebas estadísticas proporcionen valores de significación válidos. Por último, permite el uso de placebo y técnicas de enmascaramiento.

El hecho de que la asignación aleatoria proporcione grupos comparables depende, en la práctica, de dos aspectos muy relacionados: por un lado, debe generarse una secuencia aleatoria que no sea predecible ni por el investigador ni por el paciente, y por otro, es esencial que la secuencia se aplique después de haber tomado la decisión de incluir al sujeto en el estudio, de forma que ésta no pueda verse influida por el conocimiento previo de la intervención que debería recibir (ocultación de la secuencia aleatoria).

La ocultación de la secuencia aleatoria no debe confundirse con el ciego o enmascaramiento. La ocultación de la secuencia aleatoria trata de prevenir un sesgo de selección evitando que el investigador conozca la secuencia antes de la asignación, mientras que las técnicas de ciego tratan de evitar un sesgo de información y protegen la aleatorización una vez se ha producido la asignación.

**Ejemplo 17.3.** Chalmers et al (1983) clasificaron 145 ensayos clínicos sobre el tratamiento del infarto

agudo de miocardio en tres grupos. El primero de ellos correspondía a aquellos en que el investigador había sido informado del tratamiento que debía asignar a un paciente después de haber decidido incluirlo en el estudio (ocultación de la secuencia de aleatorización). El segundo grupo lo formaban los estudios en que el tratamiento asignado había sido conocido antes de incluir al paciente (asignación aleatoria no oculta), y el tercero, aquellos en los que la asignación no se había realizado de forma aleatoria. El porcentaje de variables pronósticas que no presentaron una distribución similar en los grupos de estudio fue del 35,7 y el 34,4%, respectivamente. Sólo en el primer grupo este porcentaje se mantenía por debajo del 5% esperable por azar. En los otros dos grupos se encontraron diferencias en dichas distribuciones que favorecían al grupo de tratamiento en más del 75% de los casos, y conducían a una mayor frecuencia de observación de resultados significativos a favor del tratamiento ensayado, lo que sugiere un sesgo debido al investigador.

**Ejemplo 17.4.** Odgaard-Jensen et al (2011) realizaron una revisión sistemática con la finalidad de evaluar los efectos de la asignación aleatoria y la ocultación de la secuencia de asignación. Incluyeron 10 estudios que comparaban ensayos aleatorios y no aleatorios, y encontraron que tres de ellos mostraban estimaciones similares del efecto, mientras que en dos las estimaciones eran superiores en los ensayos aleatorios, y en otros dos en los no aleatorios, y el resto eran no concluyentes. También identificaron 9 estudios que comparaban los ensayos con y sin ocultación de la secuencia de asignación, de los que 5 mostraban estimaciones superiores en los ensayos sin ocultación y 4 no mostraban diferencias significativas. Los autores concluyen que los resultados de los ensayos pueden diferir según estas características metodológicas, sin que generalmente pueda predecirse la magnitud, o incluso la dirección, de las distorsiones que se producen en la estimación del efecto.

Es importante asegurarse de que la asignación se realiza de forma correcta. Una técnica muy frecuente es la utilización de sobres cerrados numerados. Cuando un paciente cumple los criterios de selección y se decide su inclusión en el estudio, se le asigna un código (habitualmente un número secuencial) y se abre el sobre correspondiente para conocer qué intervención le ha correspondido. Si el estudio se efectúa con la técnica de doble ciego, el sobre contiene un número que codifica la intervención o el tratamiento correspondiente. En los estudios multicéntricos suele utilizarse un procedimiento

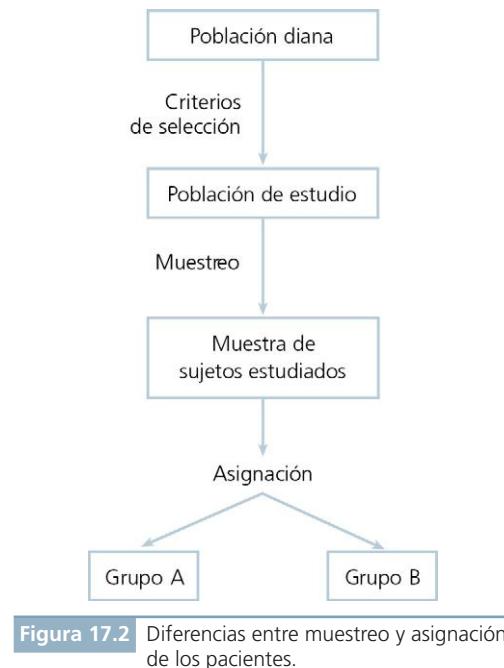
telefónico centralizado con el que contactan los investigadores cuando incluyen un nuevo paciente.

Es conveniente que la mecánica del proceso de asignación recaiga en personas o servicios no directamente involucrados en el estudio. Por ello, cada vez será más frecuente el uso de procesos automatizados para realizar la asignación aleatoria.

## Técnicas de asignación

Aunque se utiliza una terminología similar, muestreo y asignación son dos conceptos diferentes (fig. 17.2). La asignación parte de una muestra de pacientes que debe ser distribuida en dos o más grupos. En algunas ocasiones, existe un registro previo de la población candidata, a partir del cual puede obtenerse una muestra aleatoria. Después, se procede a la distribución aleatoria de los sujetos de esta muestra en dos grupos, cada uno de los cuales recibirá una de las intervenciones que se comparan. Si no se conocen previamente los sujetos candidatos, deberá prepararse una secuencia aleatoria de tratamientos, de forma que los pacientes sean asignados a partir de esta secuencia a medida que vayan siendo incluidos en el estudio.

Una forma sencilla de proceder a la distribución de los sujetos en dos grupos es la *asignación sistemática*, según la cual se asignan alternativamente según la secuencia ABABABAB... Sin embargo, este



método no es aleatorio, salvo en todo caso para el primer individuo, y no es aconsejable su uso, ya que el investigador conoce qué tratamiento recibirá el próximo sujeto que entre en el estudio, lo que puede, por ejemplo, influir en su decisión de incluirlo o no. Lo mismo ocurre si la asignación se realiza según las fechas de nacimiento (días pares o impares) o las iniciales de los apellidos.

La mayoría de las técnicas aleatorias asignan los sujetos según una probabilidad prefijada, habitualmente la misma para todos los grupos, y que no se modifica a lo largo del estudio. A continuación se presentan los métodos utilizados con mayor frecuencia (cuadro 17.3).

### Asignación aleatoria simple

Es la técnica más sencilla (cuadro 17.4). Sería el equivalente a lanzar una moneda al aire cada vez que un sujeto fuese incluido en el estudio, de forma que si saliera cara recibiría la intervención A y si saliera cruz, la intervención B. Con este método, aproximadamente la mitad de los individuos recibiría cada una de las intervenciones.

Cuadro 17.3 Técnicas de asignación de los sujetos a los grupos de estudio

#### Sistématica

#### Aleatoria

- Con probabilidad prefijada
  - Asignación aleatoria simple
  - Asignación por bloques
  - Asignación estratificada
- Adaptativa

Cuadro 17.4 Ventajas y desventajas de la asignación aleatoria simple

#### Ventajas

- Técnica muy sencilla

#### Desventajas

- Puede producir desequilibrios en el número de pacientes asignados a cada uno de los grupos, sólo por azar
- Puede producir desequilibrios en el número de sujetos asignados a cada uno de los grupos si el estudio finaliza antes de haber incluido a todos los sujetos previstos

Los ordenadores y muchas calculadoras pueden generar números aleatorios que van de 0 a 0,9999. Si se trata de asignar los sujetos a dos grupos, puede establecerse, por ejemplo, que si se obtiene para un individuo un número entre 0 y 0,4999, recibirá la intervención A, mientras que si se obtiene un número entre 0,5 y 0,9999, recibirá la B. Este método se puede generalizar cuando son más de dos grupos y también permite utilizar probabilidades de asignación diferentes para cada uno de los grupos.

**Ejemplo 17.5.** En un ensayo clínico se desea comparar tres grupos, con unas probabilidades de asignación del 25, el 25 y el 50%, respectivamente. Puede establecerse una correspondencia como la siguiente:

- Entre 0 y 0,2499 = A
- Entre 0,25 y 0,4999 = B
- Entre 0,5 y 0,9999 = C

Si se estudia a un número reducido de pacientes, puede existir un cierto desequilibrio en la cantidad de ellos que ha sido asignada a cada uno de los grupos. Existen dos técnicas para prevenir este problema: supongamos que se dispone de una muestra de 40 sujetos, y se desea distribuirla en dos grupos, de forma que cada uno tenga el mismo número de pacientes. Se procede a la lectura de la tabla de números aleatorios y se obtiene una secuencia de 20 números de dos cifras entre 0 y 40, que corresponden a los sujetos que recibirán el tratamiento A. Los 20 individuos restantes recibirán el tratamiento B. Sin embargo, esta técnica no previene la aparición de desequilibrios si se finaliza el estudio antes de lo previsto, ya que la distribución se equilibra al asignar los 40 sujetos. La otra técnica es la asignación por bloques.

### Asignación por bloques

Esta técnica consiste en crear unos *bloques* de un tamaño prefijado, múltiplo de dos, de forma que la mitad de los sujetos de cada uno de estos bloques será asignada a un grupo, y la otra mitad, al otro. El orden en que estas intervenciones se asignan dentro de los bloques es aleatorio, y el proceso se repite sucesivamente para cada uno de ellos, hasta que todos los sujetos han sido asignados. Esta técnica garantiza que, en cualquier momento a lo largo del proceso, el número de sujetos de cada uno de los grupos será muy similar.

**Ejemplo 17.6.** Supongamos que se pretende estudiar una muestra de 60 individuos, a los que se desea asignar aleatoriamente a dos tratamientos, A y B.

Si se utilizara la técnica de la asignación aleatoria simple, podría ocurrir que, si sólo se reclutaran 42 pacientes, 27 hubieran recibido el tratamiento A, y los 15 restantes, el B. La asignación por bloques previene este desequilibrio. Pueden utilizarse, por ejemplo, bloques de 6 individuos. Se establece una correspondencia entre los números aleatorios y las posibles combinaciones de las asignaciones en los bloques. Existen 20 permutaciones de bloques de 6 elementos (**tabla 17.1**), cada una de las cuales se corresponde con 5 números aleatorios de 2 cifras (del 00 al 99).

Dado que se pretende estudiar a 60 individuos y que los bloques son de 6 sujetos, se precisarán 10 bloques. A partir de una tabla se seleccionan 10 números aleatorios de dos cifras. Utilizando las correspondencias de la **tabla 17.1**, se sustituye cada uno de ellos por la permutación de 6 pacientes correspondiente. Si los números obtenidos son 21, 47, 29, etc., la secuencia de asignación de los tratamientos sería ABAABB ABBBAA ABABAB... Como puede apreciarse, en ningún momento existirá una diferencia de más de tres sujetos (la mitad del tamaño del bloque) entre ambos grupos.

Otra ventaja de este procedimiento es que, si el tipo de sujetos reclutados cambia durante el período de inclusión (p. ej., porque se utilizan consecutivamente diferentes fuentes de pacientes), la asignación por bloques producirá grupos que incluso serán más comparables.

## Asignación estratificada

Implica la división de la muestra en subgrupos en función de las variables pronósticas consideradas, y a continuación se procede a la asignación aleatoria de los sujetos dentro de cada uno de los estratos, utilizando alguna de las técnicas ya comentadas. Esta técnica ayuda a prevenir la aparición de desequilibrios en la distribución de las variables pronósticas (**cuadro 17.5**).

Cuadro 17.5 **Ventajas y desventajas de la asignación aleatoria estratificada**

### Ventajas

- Previene desequilibrios en la distribución de las variables pronósticas consideradas en la estratificación (muy útil en estudios con un reducido número de sujetos)
- Disminuye la variabilidad dentro de los estratos, y aumenta, en consecuencia, la potencia estadística del estudio

### Desventajas

- Complica el análisis estadístico
- Si existen muchos estratos, puede haber un número muy reducido de sujetos en alguno de ellos (no se aconsejan más de 8-10 estratos)

**Tabla 17.1. Asignación por bloques. Permutaciones de las asignaciones de dos tratamientos en bloques de 6 sujetos**

Permutación	Números aleatorios	Permutación	Números aleatorios
AAABBB	00-04	BAAABB	50-54
AABABB	05-09	BAABAB	55-59
AABBAB	10-14	BAABBA	60-64
AABBBA	15-19	BABAAB	65-69
ABAABB	20-24	BABABA	70-74
ABABAB	25-29	BABBAA	75-79
ABABBA	30-34	BBAAAB	80-84
ABBAAB	35-39	BBAABA	85-89
ABBABA	40-44	BBABAA	90-94
ABBBAA	45-49	BBBBAA	95-99

A partir de una tabla de dígitos aleatorios, se seleccionan números de dos cifras, cada uno de los cuales corresponde a un bloque de 6 pacientes, a los que se asignan los dos tratamientos en la secuencia correspondiente.

**Ejemplo 17.7.** Supongamos que en un estudio sobre la eficacia de dos tratamientos, A y B, para la hipertensión arterial, se incluyen en un ensayo clínico aleatorio 80 pacientes hipertensos esenciales con edades comprendidas entre los 40 y los 70 años. Se considera que la edad y la presencia de obesidad son variables pronósticas que pueden influir en la respuesta a dichos tratamientos. Dado que el tamaño de la muestra no es muy elevado, se desea prevenir la posible aparición de desequilibrios en la distribución de estas variables en los grupos de estudio. Para ello se procede a una asignación estratificada (tabla 17.2). En primer lugar, se divide la muestra en función de tres grupos de edad (40-49, 50-59 y 60-69 años, respectivamente) y la presencia o ausencia de obesidad, y se distribuye a los 80 individuos en los 6 estratos. Posteriormente, se procede a la asignación de los sujetos que forman cada uno de los estratos a los dos grupos de tratamiento, ya sea con la técnica aleatoria simple o por bloques.

El número de estratos que se crean es igual al producto del número de subgrupos que se consideran por cada variable de estratificación. Si se utilizan muchas variables, se crea un gran número de estratos, lo que hace que el tamaño de alguno de ellos pueda ser muy reducido. No se aconseja crear más de 8 o 10 estratos.

Las variables de estratificación son diferentes para cada estudio y deben considerarse únicamente las que tengan verdadera importancia sobre la respuesta a la intervención. No tiene sentido estratificar por variables fisiológicas o sociodemográficas, como la edad, el sexo o la profesión, si no existe ningún motivo para sospechar que puedan modificar la respuesta o su evaluación. También debe tenerse presente que los estratos han de ser excluyentes para evitar que existan ambigüedades en la inclusión de los sujetos a los diferentes subgrupos. Al mismo tiempo, estos estratos deben ser lo suficientemente

distintos como para sospechar que los resultados puedan ser diferentes en cada uno de ellos, y homogéneos en su interior para facilitar la detección de diferencias.

Dentro de cada estrato es preferible usar una técnica de asignación por bloques, con el fin de que el proceso sea más efectivo.

Una asignación estratificada debe tenerse en cuenta en el momento del análisis. Los individuos incluidos en cada uno de los estratos son similares en cuanto a las características pronósticas consideradas, por lo que la variabilidad es menor y permite aumentar la potencia del ensayo si se utilizan las técnicas estadísticas adecuadas.

## Técnicas adaptativas

Existen otras técnicas en las que la probabilidad de asignación se va modificando, adaptándose, a medida que el estudio progresá. Pueden clasificarse en dos grandes grupos. Por un lado, las que adaptan la probabilidad de asignación según la aparición de desequilibrios en la distribución de las variables pronósticas o en el número de sujetos de cada uno de los grupos. Por otro, las que la ajustan en función de la respuesta observada en los pacientes estudiados, buscando asignar un mayor número de pacientes al grupo que recibe el tratamiento más eficaz. La descripción de estas técnicas sobrepasa el ámbito de este texto.

## Asignación por grupos

En algunas ocasiones resulta más práctico asignar grupos de pacientes, por ejemplo, por centros, por consultas o por áreas geográficas. Este procedimiento es útil para evitar el fenómeno de *contaminación* que aparece cuando un individuo se somete, total o parcialmente, a la intervención destinada a los sujetos del otro grupo.

**Tabla 17.2. Asignación estratificada según la edad y la obesidad, con posterior asignación en bloques de 4 sujetos en cada uno de los estratos (ejemplo 17.6)**

Edad (años)	Obesidad	Número de sujetos	Asignación
40-49	Sí	8	ABAB BBAA
	No	12	BBAA ABBA...
50-59	Sí	12	AABB BABA...
	No	8	AABB ABBA
60-69	Sí	28	BAAB ABAB...
	No	12	BBAA ABBA...

**Ejemplo 17.8.** En el *Edinburgh Randomised Trial of Breast Cancer Screening*, cuyo objetivo era determinar si el cribado reducía la mortalidad por cáncer de mama, los médicos consideraron inaceptable ofrecer el cribado a unas pacientes y no ofrecerlo a las otras. Además, existía el riesgo de que las pacientes del grupo control conocieran la existencia del cribado en la consulta y lo solicitaran, lo que hubiera supuesto una contaminación. Por tanto, se decidió utilizar la consulta del médico general como unidad de asignación. De las 78 consultas que se incluyeron, 37 fueron asignadas al grupo de cribado, y las 41 restantes al grupo control. Se estudiaron 45.130 mujeres residentes en la ciudad, con edades comprendidas entre 45 y 64 años, que acudieron a dichas consultas durante un período de 4 años. La distribución por edades de estas pacientes fue prácticamente idéntica en ambos grupos. Dado que la consulta es la unidad de asignación, el análisis requirió la comparación de las tasas de mortalidad por cáncer de mama de cada una de ellas. Sin embargo, el estado socioeconómico, muy relacionado con el tipo de consulta, presentó grandes diferencias en su distribución, con una proporción

claramente mayor de pacientes de clase socioeconómica elevada en el grupo de cribado respecto al control (el 52,5 y el 15,3%, respectivamente). Ello obligó a los autores a tener en consideración esta variable en el análisis, para evitar su posible efecto de confusión (Alexander et al, 1989).

Si los integrantes de los grupos presentan alguna característica común que influya en los resultados, pequeñas diferencias en la distribución de las unidades pueden comportar grandes diferencias en la estimación de la respuesta. En el ejemplo anterior puede apreciarse que sólo existen 78 unidades de asignación, cada una de ellas integrada por un promedio de casi 6.000 mujeres. Así pues, un desequilibrio de tan sólo 3 o 4 consultas en las que predominen mujeres de clase socioeconómica elevada implica un desequilibrio de miles de mujeres entre los grupos.

Otras veces no es factible realizar la asignación individual. Es habitual el uso de la asignación por grupos en los estudios de fluoración de las aguas, en los que la unidad de asignación es la comunidad, o en los ensayos de intervención dietética, en los que lo es la familia.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- Alexander F, Roberts MM, Lutz W, Hepburn W. Randomization by cluster and the problem of social class bias. *J Epidemiol Community Health*. 1989;43:29-36.
- Ludvigsson JF, Wahlstrom J, Grunewald J, Ekblom A, Montgomery SM. Coeliac disease and risk of tuberculosis: a population based cohort study. *Thorax*. 2007;62:23-8.
- Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled clinical trials. *N Engl J Med*. 1983;309:1358-61.
- Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H, et al. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev*. 2011;(4). MR000012.
- Pomp ER, Van Stralen KJ, Le Cessie S, Vandebroucke JP, Rosendaal FR, Doggen CJ. Experience with multiple control groups in a large population-based case-control study on genetic and environmental risk factors. *Eur J Epidemiol*. 2010;25:459-66.

## BIBLIOGRAFÍA

- Altman DG, Bland JM. Treatment allocation in controlled trials: why randomise? *BMJ*. 1999;318:1209.
- Altman DG, Schulz KF. Concealing treatment allocation in randomised trials. *BMJ*. 2001;323:446-7.
- Bracken MB. On stratification, minimization and protection against types 1 and 2 error. *J Clin Epidemiol*. 2001;54:104-5.
- Chalmers I. Comparing like with like: some historical milestones in the evolution of methods to create unbiased comparison groups in therapeutic experiments. *Int J Epidemiol*. 2001;30: 1156-64.
- Cobo Valeri E. Necesidad y limitaciones de la asignación aleatoria. *Med Clin (Barc)*. 2000;115:73-7.
- de Graaf MA, Jager KJ, Zoccali C, Dekker FW. Matching, an appealing method to avoid confounding? *Nephron Clin Pract*. 2011;118:c315-8.
- Doig GS, Simpson F. Randomization and allocation concealment: a practical guide for researchers. *J Crit Care*. 2005;20:187-91.
- Dumville JC, Hahn S, Miles JN, Torgerson DJ. The use of

- unequal randomisation ratios in clinical trials: a review. *Contemp Clin Trials.* 2006;27:1-12.
- Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. *J Clin Epidemiol.* 1999;52:19-26.
- La Caze A, Djulbegovic B, Senn S. What does randomisation achieve? *Evid Based Med.* 2012;17(1):1-2.
- le Cessie S, Nagelkerke N, Rosendaal FR, van Stralen KJ, Pomp ER, van Houwelingen HC. Combining matched and unmatched control groups in case-control studies. *Am J Epidemiol.* 2008;168:1204-10.
- Lipchik GL, Nicholson RA, Penzien DB. Allocation of patients to conditions in headache clinical trials: randomization, stratification, and treatment matching. *Headache.* 2005;45:419-28.
- Puffer S, Torgerson DJ, Watson J. Cluster randomized controlled trials. *J Eval Clin Pract.* 2005;11:479-83.
- Rose S, van der Laan MJ. Why Match? Investigating Matched Case-Control Study Designs with Causal Effect Estimation. *Int J Biostat.* 2009;5. Article1.
- Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA.* 1995;273: 408-12.
- Schulz KF, Altman DG, Moher D. Allocation concealment in clinical trials. *JAMA.* 2002;288:2406-7.
- Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet.* 2002;359: 515-9.
- Schulz KF, Grimes DA. Unequal group sizes in randomised trials: guarding against guessing. *Lancet.* 2002;359:966-70.
- Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials: a review. *Control Clin Trials.* 2002;23:662-74.
- Stanley K. Design of randomized controlled trials. *Circulation.* 2007;115:1164-9.
- Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *BMJ.* 2001;322:355-7.
- Viera AJ, Bangdiwala SI. Eliminating bias in randomized controlled trials: importance of allocation concealment and masking. *Fam Med.* 2007;39:132-7.

## Medición de variables

Medir es asignar valores a una variable. Algunas variables son sencillas de medir. La sensibilidad de un microorganismo a un antibiótico o el peso de una persona no presentan grandes dificultades de conceptualización ni de medición. Estas variables se denominan *objetivas*, ya que su definición es clara, deja poco margen a la interpretación por parte del sujeto o del investigador y se dispone de un buen instrumento para medirlas. Sin embargo, muchos de los problemas que se investigan en ciencias de la salud son difíciles de medir. La gravedad de la enfermedad, la intensidad del dolor, o el propio concepto de salud o de calidad de vida son variables *subjetivas* para las que no existe una definición ni un proceso de medición universalmente aceptados.

En cualquier estudio, el equipo investigador debe asegurar la validez y la fiabilidad de las medidas que utiliza (fig. 18.1). La *validez* expresa el grado en que el valor que se obtiene con el proceso de medición se corresponde con el verdadero valor de la variable en un sujeto, es decir, que no existe un sesgo (error sistemático) en el proceso de medición.

**Ejemplo 18.1.** Supongamos que se pretende estimar la prevalencia de obesidad en una comunidad. Para ello, se pesa a una muestra de sujetos y se registra para cada uno de ellos si es obeso o no, considerando obesa aquella persona cuyo peso supere los 75 kg. Este proceso de medición de la obesidad carece de validez, ya que pueden existir sujetos que pesen menos de 75 kg y sean obesos, aunque de baja estatura, y otros que pesen más de 75 kg pero que no sean obesos por ser de mayor estatura.

La *fiabilidad* expresa el grado en que se obtendrían valores similares al aplicar el proceso de medición

en más de una ocasión, concepto muy relacionado con el de *variabilidad aleatoria*.

**Ejemplo 18.2.** Supongamos que se utiliza una tira reactiva colorimétrica para determinar la intensidad de una proteinuria en una muestra de 20 sujetos. Dos observadores independientes las evalúan y coinciden tan sólo en cinco de ellas. La determinación de la proteinuria por este método no es fiable por la gran variabilidad hallada entre los observadores.

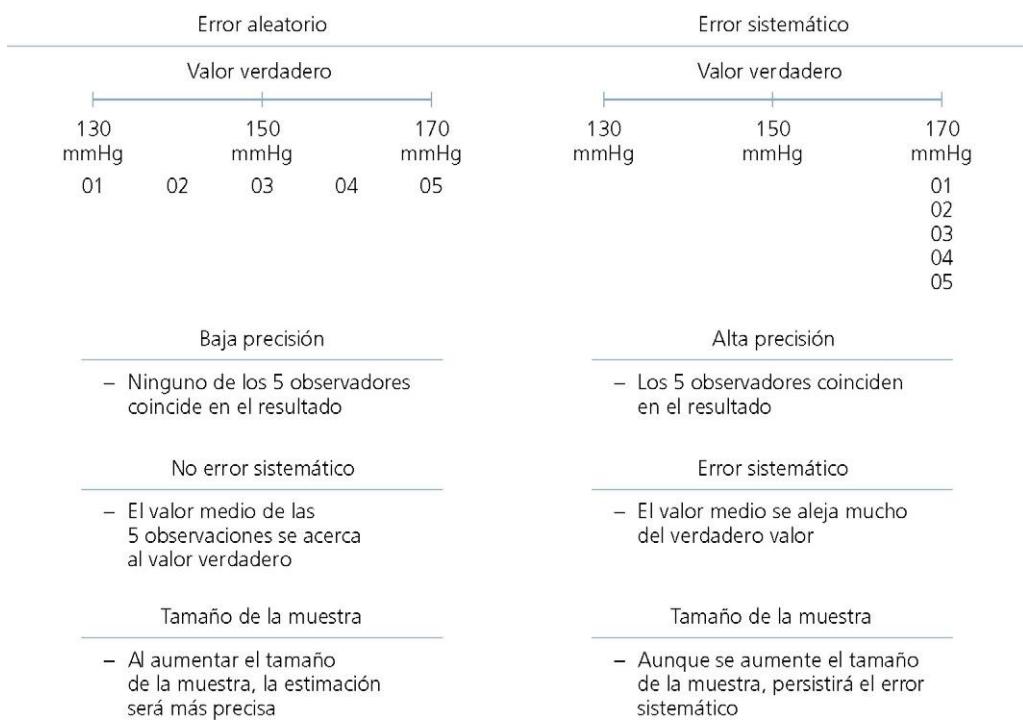
### FIABILIDAD

El término *fiabilidad* es sinónimo de precisión. Si, por ejemplo, el resultado de la medición de la concentración sérica de colesterol en un sujeto es de 180 mg/dl y, acto seguido, una segunda medición efectuada con la misma técnica y en el mismo individuo, proporciona el resultado de 240 mg/dl, de inmediato se creerá que el proceso de medición es poco fiable. Cuanta menor variabilidad se obtenga al repetir una medición, más fiable será el proceso.

El que una medida sea muy precisa no significa que tenga que ser válida. Un ejemplo sencillo sería pesar dos veces consecutivas a un mismo individuo en una báscula mal calibrada; seguramente se obtendrían resultados similares, pero no válidos, ya que la balanza proporcionaría mediciones inexactas.

### Fuentes de variabilidad

La *fiabilidad* de una medida no es una propiedad intrínseca de ella, sino más bien de un instrumento cuando se aplica en una muestra de individuos concreta, en unas condiciones dadas. La medición de un



**Figura 18.1** Relación entre error aleatorio y error sistemático. Medición de la presión arterial sistólica de un paciente por cinco observadores independientes (01, 02, 03, 04 y 05).

fenómeno clínico está sujeta a la variación biológica individual, del observador o del instrumento utilizado. Así, si dos médicos toman la presión arterial de un paciente y no obtienen el mismo resultado, puede deberse a la propia variación de la tensión arterial entre una medición y otra, a que el esfigmomanómetro utilizado proporciona resultados variables, a que los clínicos difieren en la forma de medir y registrar la presión arterial, o a una combinación de todas estas circunstancias. Aunque se describirá cada una de las fuentes de variación por separado, en la práctica todas ellas actúan de forma simultánea.

### Variación individual

La mayoría de los fenómenos biológicos varían de un momento a otro. Una medición realizada en un momento dado puede considerarse una muestra de todas las posibles mediciones de este fenómeno durante un período determinado y no tiene por qué representar su verdadero valor.

**Ejemplo 18.3.** Las variaciones del colesterol sérico en un mismo individuo pueden ser lo suficientemente grandes como para que sea difícil evaluar su riesgo

cardiovascular en una sola determinación. Para poner de relieve este hecho se realizó un trabajo en el que se analizaba el colesterol sérico de un individuo en ayunas cada 20 minutos durante 6 horas (Natelson et al., 1988). Los valores de colesterol variaron desde un mínimo de 189 mg/dl hasta un máximo de 211 mg/dl. De las 19 muestras extraídas, 8 presentaban valores por encima de 200 mg/dl. Por consiguiente, si el criterio de inclusión en un estudio fuera precisamente tener unos valores superiores a 200 mg/dl, esta persona hubiera podido o no ser candidata, según la muestra considerada. Igualmente, es posible que un individuo presente valores más bajos o más altos de colesterol una vez finalizado un estudio, con independencia de la intervención practicada.

**Ejemplo 18.4.** En un estudio en el que se analizaban los registros de 8 tomas de la presión arterial de 30 individuos se observó una gran variabilidad intra-individuo. La variabilidad media de la presión sistólica fue de 13,2 mmHg (desviación estándar [DE]: 6,8) y de la presión diastólica de 8,8 mmHg (DE: 4,6) utilizando un esfigmomanómetro de mercurio (Cuesta y Gómez Moro, 1987). Los autores

observaron también una diferencia significativa entre la presión sistólica del brazo derecho respecto a la observada simultáneamente en el brazo izquierdo. Las cifras diastólicas no mostraron diferencias de uno a otro brazo.

Para minimizar la variación biológica intraindividual, una estrategia consiste en repetir la medición varias veces y utilizar alguna medida promedio de los resultados. Sin embargo, antes de ponerla en marcha hay que considerar el coste y las dificultades prácticas que implica.

### Variación causada por el instrumento

Por *instrumento* se entiende cualquier vehículo útil para recoger datos de una forma organizada; por ejemplo, un esfigmomanómetro, un autoanalizador o un cuestionario. Siempre pueden presentarse variaciones al aplicar un instrumento de medida y en la técnica empleada. Por ejemplo, al tomar la presión arterial, la presión del brazal o la posición del estetoscopio pueden cambiar, lo que originará inconsistencias en las mediciones. Por ello, es importante estandarizar los procedimientos y seleccionar el instrumento que presente menor variabilidad.

### Variación causada por el observador

La tercera fuente de variabilidad es el observador. Las personas que llevan a cabo la medición de la presión arterial pueden variar; por ejemplo, el ángulo de observación, las condiciones en que realizan las mediciones, o aumentar su experiencia a medida que prograza el estudio.

Para reducir la variabilidad debida a los observadores es necesario entrenarlos debidamente antes de iniciar el estudio y elegir medidas lo más objetivas posible, ya que, cuanto más subjetivas sean, mayor es la probabilidad de que exista variabilidad de una medición a otra y entre los distintos observadores. Por último, siempre que sea factible, es conveniente evaluar la variabilidad interobservador e intraobservador.

### Evaluación de la fiabilidad

La fiabilidad se evalúa repitiendo el proceso de medición, con el fin de analizar la concordancia entre las distintas medidas. El índice más utilizado para ello, en caso de variables cualitativas, es el índice kappa, que corrige la concordancia observada por la que se esperaría simplemente por azar (anexo 4). Si las variables son cuantitativas, se utiliza el coeficiente de correlación intraclass, que combina un

análisis de la correlación de las medidas con una prueba de la diferencia de las medias entre ellas, o el método gráfico de Bland y Altman (anexo 4).

Deben estudiarse diferentes aspectos de la fiabilidad: la repetibilidad (fiabilidad test-retest), la concordancia intraobservador y la concordancia interobservador.

### Repetibilidad

La evaluación de la repetibilidad tiene por objetivo determinar si una prueba da los mismos resultados o similares cuando se aplica a una misma persona en más de una ocasión. Requiere que las condiciones de aplicación sean iguales en ambas ocasiones.

El hecho de aplicar la misma prueba en más de una ocasión en la misma muestra de individuos conlleva algunos problemas, ya que algunas características pueden variar con el tiempo. Es el caso de una medida sobre el estado físico o de salud de una persona; cuanto más tiempo transcurra entre ambas pruebas, mayor es la probabilidad de que hayan ocurrido cambios reales en su estado de salud, o de que su opinión o actitud hacia un problema de salud haya cambiado. En esta situación, la fiabilidad de la prueba será infravalorada. Esta dificultad se puede subsanar, al menos parcialmente, efectuando la segunda medición poco tiempo después de la primera. Sin embargo, si el intervalo de tiempo entre la aplicación de las dos pruebas es breve, es posible que el resultado obtenido en la segunda ocasión esté influido por el aprendizaje adquirido la primera vez. En otras palabras, ambas mediciones no son independientes y el coeficiente de fiabilidad estará artificialmente elevado.

Cabe la posibilidad de que los participantes no accedan a que se les efectúe la misma prueba más de una vez en poco tiempo, en especial cuando sea dolorosa o molesta, o implique mucho tiempo de realización o desplazamiento, lo cual significa que se perderá mucha información, ya que si un individuo no ha completado las dos pruebas no será útil para el análisis.

### Concordancia intraobservador e interobservador

La *concordancia intraobservador* se refiere al grado de consistencia de un observador consigo mismo al leer o interpretar un resultado. Al igual que con la repetibilidad, las dos medidas que realiza cada observador deben ser independientes una de otra, lo que puede ser muy difícil de conseguir cuando la observación de interés es un hallazgo físico que requiere la presencia directa del paciente, ya que el

recuerdo de la primera exploración puede condicionar el resultado de la segunda. Esta dificultad queda soslayada si la observación es, por ejemplo, una radiografía o un electrocardiograma.

Por *concordancia interobservador* se entiende la consistencia entre dos observadores independientes sobre una medida practicada en el mismo individuo. Para asegurar la independencia de las medidas entre ambos observadores, uno no debe conocer el resultado proporcionado por el otro (observadores ciegos).

Con frecuencia, la concordancia intraobservador e interobservador se evalúa en un mismo estudio. En este caso, hay que asegurar la independencia de todas las medidas, para lo que puede ser muy útil aplicar técnicas como la aleatoriedad en la secuencia de aplicación de las medidas y las técnicas de enmascaramiento.

El análisis de la concordancia interobservador e intraobservador es, en muchas ocasiones, un paso previo a la validación de una nueva prueba diagnóstica, ya que, si los distintos observadores discrepan en los resultados, la prueba, aunque teóricamente pueda ser válida, tendrá poca utilidad clínica.

## Consecuencias de una medida poco fiable

La fiabilidad de las medidas no se ha de analizar en cada estudio. Sólo es necesario cuando se desarrolle una nueva medida o como control de la calidad de las mediciones. Sin embargo, al planificar un trabajo de investigación es imprescindible considerar las posibles medidas alternativas de las distintas variables y escoger las que hayan demostrado ser más fiables en estudios anteriores.

Si la medición de la variable de respuesta tiene poca fiabilidad, la consecuencia directa es que la estimación del efecto que se obtenga será poco precisa, es decir, su intervalo de confianza será muy amplio. Este problema puede soslayarse aumentando el tamaño de la muestra, ya que el número de sujetos necesario depende, entre otros factores, de la variabilidad de las medidas, aunque ello supone un incremento de la complejidad y el coste del estudio.

## VALIDEZ

El término validez se refiere al grado en que una variable mide realmente aquello para lo que está destinada. Es un concepto ligado al de error sistemático. Cuanto menos válida sea una medida, más probabilidades hay de cometer un sesgo.

Unas variables son más válidas que otras. Por ejemplo, la hemoglobina glucosilada representa mejor el grado de control de un diabético que una medición aislada de la glucemia. Los responsables del estudio deben procurar escoger siempre las medidas más válidas, en especial cuando se trata de las variables importantes del estudio.

## Fuentes de error sistemático

Las posibles fuentes de error sistemático son las mismas que las enumeradas en el apartado destinado a la fiabilidad.

### Error causado por el individuo

Ocurre cuando el error en la medición de un fenómeno se introduce a causa de los participantes en el estudio. El ejemplo más sencillo es el sesgo de memoria. Los pacientes que sospechen que el problema de salud que padecen está relacionado con alguna variable en estudio, es posible que recuerden con mucha más exactitud su historia pasada de exposición que aquellos participantes libres de la enfermedad. Este diferente recuerdo se traduce en un error en la medición de la variable en estudio.

**Ejemplo 18.5.** Supongamos que para investigar los posibles efectos teratogénicos de un fármaco se realiza un estudio de casos y controles. Los casos son los recién nacidos con malformaciones, y los controles, recién nacidos sin malformaciones. Si se pregunta a las madres por el consumo de fármacos durante el embarazo, es muy probable que las madres de niños con malformaciones tiendan a recordar con más detalle todo lo ocurrido durante su embarazo, incluidos los fármacos que tomaron, que las madres de niños sin malformaciones.

En otras ocasiones, los participantes pueden falsopear de forma deliberada las respuestas simplemente para complacer al entrevistador o porque piensen que una determinada respuesta no está bien vista socialmente, por ejemplo, cuando se interroga sobre los hábitos tóxicos de una persona.

### Error causado por el observador

Ocurre cuando el observador mide una variable de forma errónea, por ejemplo, al utilizar una técnica incorrecta para la toma de la presión arterial (mala posición del brazal, tomarla en el brazo inadecuado, etc.) o al administrar un cuestionario de forma errónea.

En los estudios analíticos puede ocurrir también cuando el observador mide sistemáticamente una variable de forma distinta en un individuo u otro, según el grupo de estudio al que pertenece. Si un investigador supone que los pacientes que reciben un tratamiento presentan con más frecuencia un determinado efecto secundario, es posible que, incluso de forma inconsciente, lo busque con mayor insistencia (con preguntas más frecuentes o con más pruebas complementarias) en aquellos que lo reciben que en los participantes del otro grupo.

### Error causado por el instrumento

Ocurre cuando la medida empleada no es la adecuada para el fenómeno que se quiere analizar, o en el caso de que el instrumento utilizado no esté bien calibrado.

### Evaluación de la validez

Para evaluar la validez de una medida se comparan sus valores con los obtenidos con una técnica de referencia objetiva, fiable y ampliamente aceptada como una buena medida del fenómeno de interés (*validez de criterio*).

Cuando la variable es dicotómica, los índices de validez de criterio más usados son la sensibilidad y la especificidad (anexo 3). La *sensibilidad* se define como la probabilidad de que una persona que presente la característica sea clasificada correctamente por la medida empleada en el estudio. La *especificidad* se define como la probabilidad de que una persona que no tenga la característica sea correctamente clasificada por la medida utilizada. Para que una medida sea útil deberá ser lo más sensible y específica posible. Si la variable es cuantitativa, una forma de evaluar la validez es el análisis de las diferencias individuales entre los valores que proporciona la medida que se está utilizando y la obtenida con el método de referencia, sintetizadas en su media y desviación estándar (anexo 4).

A veces es difícil disponer de un criterio de referencia adecuado, bien porque no existe o porque no está al alcance del investigador. En estos casos, el procedimiento más empleado es evaluar la *validez de constructo* o *de concepto*, que analiza la correlación de la medida con otras variables que se cree que están relacionadas con ella (*validez convergente*) y la correlación con otras que se sospecha que no tienen relación alguna (*validez divergente*). Si la correlación es alta en el primer caso y baja en el segundo, puede considerarse que la medida tiene validez de constructo.

Otro aspecto que se debe tener en cuenta es la evaluación de la *validez de contenido* de la medida, es decir, si contempla o abarca todas las dimensiones del fenómeno que se quiere medir. Por ejemplo, la concentración de triglicéridos puede ser una medida fácil de definir, fiable y con cifras muy exactas por parte del laboratorio. Sin embargo, puede no ser una medida válida de arteriosclerosis. Es conveniente abordar estos problemas en la fase de diseño de un estudio, ya que después esta información ayudará a matizar los resultados.

### Consecuencias de una medida poco válida

La utilización de medidas poco válidas conduce a una respuesta errónea a la pregunta de investigación. Si se trata de un estudio descriptivo que estime, por ejemplo, la frecuencia de una enfermedad, se obtendrá una prevalencia errónea. Si se trata de estudios analíticos, los sesgos de información conducirán a una estimación errónea del efecto del factor de estudio. En este tipo de estudios conviene distinguir entre los errores diferenciales y los no diferenciales.

### Error no diferencial

Es el error que ocurre al medir el factor de estudio y/o la variable de respuesta y que se produce por igual en todos los participantes, con independencia del grupo al que pertenecen o del valor de cualquier otra variable. En otras palabras, significa que cuando la variable es cualitativa, la sensibilidad y la especificidad para medirla permanecen constantes en todos los participantes, al margen de otros factores. Si la variable es continua, lo que permanece constante es la diferencia entre la media observada y la verdadera media.

**Ejemplo 18.6.** Consideremos un estudio en el que un 60% de los individuos del grupo A desarrollan la enfermedad, mientras que en el grupo B, la incidencia es sólo del 20%. La verdadera razón de incidencias es igual a 3. Supongamos que la prueba que se aplica para detectar la enfermedad tiene una sensibilidad del 85% y una especificidad del 90% en ambos grupos. Aplicando estos valores se obtienen los datos de la tabla 18.1. Por ejemplo, en el grupo A se esperaría que de las 60 personas con la enfermedad, la prueba identificara correctamente a 51 ( $60 \times 0,85$ ), y que 36 ( $40 \times 0,90$ ) se clasifiquen correctamente como que no la tienen.

De esta forma, la incidencia acumulada en el grupo A ha disminuido del 60 al 55%, mientras que

**Tabla 18.1** Estudio hipotético que ilustra un error sistemático no diferencial (ejemplo 18.7)

Grupo A Resultados de la prueba	Enfermedad		Total
	Presente	Ausente	
Positivos	51	4	55
Negativos	9	36	45
Total	60	40	100
Incidencia acumulada verdadera: 60%.			
Sensibilidad de la prueba: 85%.			
Especificidad de la prueba: 90%.			

Grupo B Resultados de la prueba	Enfermedad		Total
	Presente	Presente	
Positivos	17	8	25
Negativos	3	72	75
Total	20	80	100
Incidencia acumulada verdadera: 20%.			Razón de incidencias verdadera: 3.
Sensibilidad de las pruebas: 85%.			Razón de incidencias observada: 2,2.
Especificidad de las pruebas: 90%.			

en el grupo B ha aumentado del 20 al 25%. Esto es así porque cuando la incidencia o la prevalencia no es muy alta, como sucede en el grupo B, incluso una pequeña proporción de falsos positivos puede compensar una mayor proporción de falsos negativos.

Si se calcula la razón de incidencias, se observa que ha disminuido de 3 a 2,2 como resultado de una mala clasificación no diferencial.

Cuando se comparan dos grupos y se usa una medida poco válida que afecta a todos los participantes por igual, se introduce un error no diferencial que conduce a una infraestimación del verdadero efecto o asociación.

Cuando el efecto real es de gran magnitud, esta infraestimación no tiene una gran trascendencia. Sin embargo, en aquellos estudios que no encuentran un efecto, el investigador debe valorar la posibilidad de que la prueba empleada sea poco sensible o específica y que ésta sea la verdadera causa de que no se haya encontrado ninguna asociación.

Muchos autores tienden a asumir que los errores que cometen en la medición de las variables son de naturaleza no diferencial, pero en ocasiones esta asunción es difícil de mantener (por ejemplo, porque existe correlación de errores entre los ítems de un cuestionario, o debido a efectos no contro-

lados de otras covariables). Hay que tener presente que pequeñas desviaciones en la asunción de independencia de los errores puede conducir a un sesgo en cualquier dirección.

### Error diferencial

Ocurre cuando la medición del factor de estudio y/o la variable de respuesta depende del valor de otra variable, y afecta de modo diferente a los participantes en el estudio según el grupo al que pertenecen.

Cuando la medida es poco válida, y la sensibilidad y la especificidad de la prueba son diferentes en cada grupo, la mala clasificación que resulta puede sesgar los resultados en cualquier dirección. Una verdadera asociación puede quedar enmascarada, disminuida o aumentada, o bien puede encontrarse un efecto que en realidad no existe.

**Ejemplo 18.8.** En un estudio similar al del ejemplo 18.7 se utiliza una prueba de detección de la enfermedad que tiene la misma especificidad en ambos grupos (98%), mientras que la sensibilidad es del 98% en el grupo A y sólo del 70% en el B. El resultado es que se sobreestima la verdadera razón de incidencias, que de 3 ha pasado a 3,75 (tabla 18.2).

**Tabla 18.2 Estudio hipotético que ilustra un error sistemático diferencial (ejemplo 18.8)**

Grupo A Resultados de la prueba	Enfermedad		Total	Incidencia observada
	Presente	Ausente		
Positivos	59	1	60	
Negativos	1	39	40	
Total	60	40	100	60%
Incidencia acumulada verdadera: 60%. Sensibilidad de la prueba: 99%. Especificidad de la prueba: 98%.				
Grupo B Resultados de la prueba	Enfermedad		Total	Incidencia observada
	Presente	Presente		
Positivos	14	2	16	
Negativos	6	78	84	
Total	20	80	100	16%
Incidencia acumulada verdadera: 20%. Sensibilidad de las pruebas: 70%. Especificidad de las pruebas: 98%.				

Un ejemplo clásico de error diferencial es el sesgo de memoria en los estudios de casos y controles, en el que el conocimiento de la enfermedad influye sobre el recuerdo de la exposición (ejemplo 18.5). Otro ejemplo son los estudios prospectivos en los que no se aplican o no es posible aplicar técnicas de ciego: el conocimiento de la exposición (o la intervención) puede condicionar la búsqueda, consciente o inconsciente, de un efecto en este individuo, de forma diferente a la realizada en otro individuo no expuesto. En ambos casos, la medición no se realiza independientemente y puede sobrevalorarse o infraestimarse la verdadera magnitud del efecto o de la asociación.

## ESTRATEGIAS PARA AUMENTAR LA FIABILIDAD Y LA VALIDEZ

Las cinco primeras estrategias tienen efecto sobre la validez y sobre la precisión; la sexta sólo mejora la fiabilidad, mientras que las restantes sirven para aumentar la validez de la medición.

1. *Seleccionar las medidas más objetivas posibles.* Si la variable de respuesta es, por ejemplo, la mortalidad total, no existirán variaciones en su determinación por parte de los observadores. Si, por el

contrario, es la aparición o el alivio de un síntoma, es más probable que la evaluación se lleve a cabo de forma dispar entre los observadores.

2. *Estandarizar la definición de las variables.* Es imprescindible que las variables estén definidas de forma operativa en el protocolo o en el manual de instrucciones, para que todos los investigadores utilicen los mismos criterios incluso ante situaciones dudosas.
3. *Formar a los observadores.* Su entrenamiento no permitirá controlar totalmente la variación entre un observador y otro, pero sí reducirla.
4. *Utilizar la mejor técnica posible.* Por ejemplo, si se tiene la posibilidad de escoger entre un aparato aneroide y otro de mercurio para medir la presión arterial, se elegirá este último, ya que es más fiable.
5. *Utilizar instrumentos automáticos.* Si es factible, y el instrumento automatizado es más preciso, permitirá reducir la variabilidad interobservador.
6. *Obtener varias mediciones de una variable.* El promedio de varias medidas tiende a ser más preciso que una medida aislada. El ejemplo más clásico es la presión arterial: dado que es una medida muy variable, la práctica más habitual es tomarla en más de una ocasión y obtener la media de las distintas mediciones. De la misma forma, se obtendrá una

estimación más precisa de un fenómeno como la calidad de vida si se realizan varias preguntas relacionadas con el tema en un cuestionario. Esta estrategia sólo mejora la precisión, no la validez.

**7. Emplear técnicas de enmascaramiento (ciego).** En los ensayos clínicos controlados con placebo donde es posible utilizar las técnicas de enmascaramiento, y más concretamente las de doble ciego, es muy difícil incurrir en un error diferencial, ya que ni los observadores ni los participantes conocen la intervención que están recibiendo. En algunos estudios observacionales es factible mantener ciegos a los observadores, pero no así a los participantes. En esta situación se elimina el error potencial que se deriva de las expectativas de los observadores, pero no el que puede provenir de

los participantes. De todos modos, las técnicas de ciego no evitan los errores no diferenciales.

- 8. Calibrar los instrumentos.** El ejemplo 18.4, correspondiente a un estudio sobre la calibración de los esfigmomanómetros, ilustra lo que puede ocurrir si no se toma esta precaución. Hay que calibrar los instrumentos con el fin de evitar medidas sesgadas.

La puesta en marcha de todas las estrategias simultáneamente es muy difícil. No obstante, la estandarización de las variables y la formación de los observadores debe realizarse siempre. La aplicación del resto de estrategias dependerá de las características de las variables que se estén estudiando, el diseño empleado y la disponibilidad de recursos.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- |  |  |  |
|--|--|--|
| Cuesta V, Gómez Moro MB. Precisión en la toma de tensión arterial: eficacia de los esfigmomanómetros | automáticos. Rev Clin Esp. 1987;180:482-5.               | cholesterol: implications for coronary prevention. Lancet. 1988;2:404-5. |
|  | Natelson B, Tapp W, Munsif A, Burns W. Fluctuating serum |  |

## BIBLIOGRAFÍA

- |   |  |   |
|---|--|---|
| Armstrong BK, White E, Saracci R. Principles of exposure measurement in epidemiology. Oxford: Oxford University Press; 1994. p. 5.  | Bland JM, Altman DG. Measurement error. BMJ. 1996;312:1654.  | the null? Int J Epidemiol. 2008;37(2):382-5.  |
| Barbara AM, Loeb M, Dolovich L, Brazil K, Russell ML. Patient self-report and medical records: measuring agreement for binary data. Can Fam Physician. 2011;57(6): 737-8.   | Bland JM, Altman DG. Measurement error proportional to the mean. BMJ. 1996;313:106.  | Kundel HL, Polansky M. Measurement of observer agreement. Radiology. 2003;228:303-8.  |
| Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet Gynecol. 2008;31(4):466-75. | Chyou PH. Patterns of bias due to differential misclassification by case-control status in a case-control study. Eur J Epidemiol. 2007;22(1):7-17.   | Latour J, Abraira V, Cabello JB, López Sánchez J. Las mediciones clínicas en cardiología: validez y errores de medición. Rev Esp Cardiol. 1997;50:117-28.       |
| Bland JM, Altman DG. Comparing two methods of clinical measurement: a personal history. Int J Epidemiol. 1995;24(Suppl 1):S7-S14.   | Greenland S, Gustafson P. Accounting for independent nondifferential misclassification does not increase certainty that an observed association is in the correct direction. Am J Epidemiol. 2006;164(1):63-8. | Pekkanen J, Sunyer J, Chinn S. Nondifferential disease misclassification may bias incidence risk ratios away from the null. J Clin Epidemiol. 2006;59(3):281-9. |
| Bland JM, Altman DG. Measurement error and correlation coefficients. BMJ. 1996;313:41-2.  | Jiménez Villa J. Comparación de métodos cuantitativos de medida. FMC. 1994;1:404-10.   | Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 2005;85(3): 257-68.                 |
|   | Jurek AM, Greenland S, Maldonado G. How far from non-differential does exposure or disease misclassification have to be to bias measures of association away from  | Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005;37(5):360-3.  |

## Selección y definición de las variables

### SELECCIÓN DE VARIABLES

La selección debe guiarse por una norma sencilla: recoger tantas variables como sean necesarias y tan pocas como sea posible. Cuando existan dudas sobre la pertinencia de incluir una variable, debe contrastarse su utilidad con la dificultad de su medición. Las variables se pueden agrupar en cinco grandes bloques (*cuadro 19.1*). Esta lista es orientativa y deberá adecuarse a cada situación concreta.

El primer grupo hace referencia a las variables que permiten evaluar la aplicabilidad del protocolo, se corresponden con los criterios de inclusión y exclusión, y sirven para determinar si un individuo es candidato para participar en el estudio. Es útil recoger y archivar esta información, ya que interesará conocer los motivos de las exclusiones para evaluar de forma adecuada la capacidad de generalización de los resultados.

El segundo grupo corresponde a dos elementos ya especificados en la formulación del objetivo. Por un lado, el factor de estudio. Si se trata de una exposición, interesarán medir el tipo, la intensidad y la duración. Si se trata de una intervención, el tipo, la dosis, la pauta y la duración. Por otro lado, la variable de respuesta, que permitirá estimar la existencia y la magnitud del efecto observado. Estas dos variables deben ser definidas y medidas con la máxima fiabilidad y validez, ya que el objetivo del estudio es cuantificar la relación entre ellas.

Las variables del tercer grupo se identifican a partir del análisis del modelo teórico en el que se enmarca la investigación. Corresponden a los potenciales factores de confusión, es decir, las variables que se sabe, o se sospecha, que están asociadas tanto al factor de estudio como a la variable de respuesta.

Cuadro 19.1 **Variables que deben ser medidas en un trabajo de investigación**

- Variables que permitan evaluar la aplicabilidad del protocolo (criterios de selección)
- Variables que permitan medir el factor o los factores de estudio, y la variable o variables de respuesta
- Variables que puedan actuar como:
  - Posibles factores de confusión
  - Posibles variables modificadoras del efecto
  - Pasos intermedios de la cadena causal
- Variables universales descriptoras de los sujetos estudiados
- Otras variables de interés (subgrupos de población, preguntas secundarias, medidas de tiempo, etc.)

Su medición permitirá controlar su efecto en el análisis. También es importante medir las variables que puedan actuar como modificadoras del efecto, para conocer y describir mejor el efecto del factor de estudio. Por último, es interesante recoger datos sobre los pasos intermedios de la cadena causal, para evaluar cómo han evolucionado y cómo se han comportado ante el resto de los factores.

El cuarto grupo lo forman las variables que describen las características de los sujetos estudiados, de forma que pueda evaluarse la capacidad de generalización de los resultados a otros grupos de sujetos o poblaciones. Son variables la clasificación de la enfermedad, el tiempo de evolución y los tratamientos previos, entre otras. También se incluyen variables que podrían llamarse universales, ya que se recogen en la mayoría de los estudios, como el sexo o la edad.

El último grupo corresponde a las variables que complementan la medición del efecto o la asociación, definen subgrupos de sujetos de especial interés o son necesarias para responder a las preguntas secundarias.

Una buena práctica es elaborar una lista de las variables contenidas en cada una de estas categorías y debatirla entre todo el equipo investigador para valorar la pertinencia de su recogida.

## DEFINICIÓN DE LAS VARIABLES

Es conveniente adoptar definiciones estándar, utilizadas y validadas por otros investigadores, con la finalidad de poder comparar los resultados con los de otros trabajos.

**Ejemplo 19.1.** En una revisión de los ensayos clínicos que evalúan las intervenciones de apoyo a los cuidadores de pacientes con demencia, Thompson y Briggs (2000) ponen de manifiesto que tanto el tipo de las intervenciones evaluadas como el de las variables de respuesta y sus definiciones son distintas en la mayoría de los estudios, lo cual dificulta enormemente tanto la comparación de los resultados como su interpretación conjunta.

Las definiciones han de ser claras, operativas y no deben dejar lugar a la ambigüedad; además, deben prever todas las situaciones posibles. Por ejemplo, en la definición de visita: ¿se incluyen las de enfermería?, ¿y las consultas telefónicas?, ¿o las debidas a motivos burocráticos?, ¿o las consultas sobre el paciente realizadas por sus familiares?

Para muchos conceptos se utilizan mediciones aproximadas. Por ejemplo, al definir la profesión de un sujeto deberá decidirse si interesa la ejercida en la actualidad o la habitual. Si se está estudiando una determinada exposición laboral, interesa más la historia ocupacional que la profesión.

A menudo no existe una variable única que, por sí sola, exprese toda la complejidad del fenómeno que se desea medir. En estas situaciones pueden utilizarse diferentes variables, de manera que cada una de ellas refleje un aspecto diferente de dicho fenómeno, y que después puedan agruparse en una escala combinada.

**Ejemplo 19.2.** Para evaluar la calidad de vida puede utilizarse un cuestionario como el SF-36 (Short-Form-36 Health Survey) validado en España (Alonso et al, 1995), que consta de 36 ítems medidos en una escala ordinal y que abarca 8 dimensiones: estado físico, limitaciones por problemas

físicos, dolor, salud mental, limitaciones de vida a causa de problemas emocionales, vitalidad, energía o fatiga y percepción de la salud general.

En ocasiones es necesario descomponer un fenómeno complejo en diferentes aspectos que se miden por variables separadas. Un ejemplo es el electrocardiograma, cuya valoración requiere estudiar el patrón QRS, el segmento ST, la onda T, el ritmo y la frecuencia, entre otros.

## ESCALAS DE MEDIDA

La escala de medida determinará el análisis estadístico que podrá realizarse. La más simple corresponde a las variables nominales, cuyos valores son categorías no numéricas bien definidas, como por ejemplo, el tipo de tratamiento de la diabetes mellitus codificado como dieta sola, dieta más hipoglucemiantes orales, dieta más insulina, y otras combinaciones. En el caso concreto de que sólo existan dos valores posibles (sí/no, presente/ausente, masculino/femenino, etc.) se habla de *variables dicotómicas*.

En las variables ordinales las categorías pueden ordenarse de alguna forma lógica. Por ejemplo, la codificación del dolor, en ausente, leve, moderado o grave.

Las variables cuantitativas discretas pueden adoptar sólo ciertos valores (en general, números enteros), como ocurre con el número de hijos o el de ingresos hospitalarios. Las variables cuantitativas continuas pueden adoptar cualquier valor numérico, en general, dentro de un rango; por ejemplo, el peso, la edad o la glucemia basal.

En muchas ocasiones, la propia definición de la variable lleva implícita la escala de medida, pero en otras puede ser necesario escoger entre diversas posibilidades. El consumo de tabaco puede expresarse como una variable nominal (sí/no, o bien nunca ha fumado/fumador/ex fumador), ordinal (no/fumador leve/fumador moderado/fumador importante) e incluso cuantitativa (número de cigarrillos o gramos de nicotina diarios). Como norma general, es preferible escoger la escala continua, si es posible. En primer lugar, porque contiene más información, lo cual permite utilizar pruebas estadísticas más potentes (la tabla 19.1 enumera las escalas por orden creciente de información que contienen). En segundo lugar, porque, a partir de los datos cuantitativos, puede agruparse como categorías cualitativas, incluso según diferentes criterios, mientras que el proceso inverso no es posible.

Sin embargo, debe valorarse la dificultad de la recogida de la información necesaria frente al

**Tabla 19.1 Escalas de medida de las variables**

Escala	Ejemplos
– Cualitativa:	
• Nominal	Sexo: masculino/femenino Cefalea: sí/no Tabaco: sí/no Vía de administración del fármaco
• Ordinal	Clase social: I, II, III, IV, V Proteinuria: -, +, ++, +++ Cefalea: no, leve, moderada, grave Tabaco: no, fumador moderado, gran fumador
– Cuantitativa:	
• Discreta	Número de hijos Número de ingresos hospitalarios Número de episodios de una enfermedad Número de visitas en el último año
• Continua	Peso Glucemia basal Consumo de alcohol (g/día) Presión arterial

beneficio esperado. Si se desea medir el consumo de alcohol, el cálculo de los gramos consumidos diariamente requiere un esfuerzo suplementario. Si esta variable tiene un interés marginal en el estudio, puede bastar con medirla con una escala ordinal (sin consumo/consumo moderado/consumo importante), mucho más fácil de obtener. Pero si la variable es importante, debe medirse con la mayor precisión y cuantitativamente, si es posible.

**Ejemplo 19.3.** Existen diversos métodos para medir la intensidad del dolor. Las escalas descriptivas simples resultan poco sensibles. Se han desarrollado otras técnicas, entre las que destacan los métodos gráficos y las escalas analógicas visuales, y estas últimas son las que se han mostrado más sensibles. Consisten en representaciones gráficas en forma de línea recta cuyos límites se definen como los límites extremos de la sensación que se quiere medir. El paciente marca en la escala el punto que, a su juicio, representa la intensidad de su síntoma. Es recomendable que la línea no contenga señales ni puntos intermedios de referencia, ya que pueden condicionar la respuesta, convirtiendo una escala inicialmente cuantitativa en una prácticamente categórica, como se ilustra en la figura 19.1. Las escalas analógicas visuales no sólo se utilizan en la valoración del dolor, sino también de otros síntomas percibidos por el paciente más o menos subjetivamente.

Una buena escala debería cumplir los criterios enumerados en el cuadro 19.2.

## FUENTES DE INFORMACIÓN

Las fuentes de obtención de datos pueden clasificarse en cinco grupos (**cuadro 19.3**). Algunas variables pueden medirse utilizando diferentes fuentes. La elección de la más adecuada se basa en el tipo y la validez de la información que pueden proporcionar, los recursos necesarios para obtenerla, la aceptabilidad del método por los sujetos y la probabilidad de que proporcione una cobertura adecuada a todos ellos.

**Ejemplo 19.4.** Un estudio comparó la información proporcionada por las mujeres respecto a la realización previa de pruebas de Papanicolaou mediante entrevista y la registrada en las historias clínicas (Walter et al., 1988). Los datos procedentes de la entrevista indicaban una mayor cantidad de pruebas realizadas en los 5 años previos, una fecha más reciente de la última exploración, una mayor presencia de sintomatología y una gran discrepancia en los resultados de la prueba. Un estudio similar comparó la información obtenida mediante entrevista sobre el consumo de anticonceptivos orales con la registrada en las historias clínicas (Rosenberg et al., 1983). En la entrevista se utilizaron recordatorios de sucesos relevantes y fotografías de las marcas disponibles en

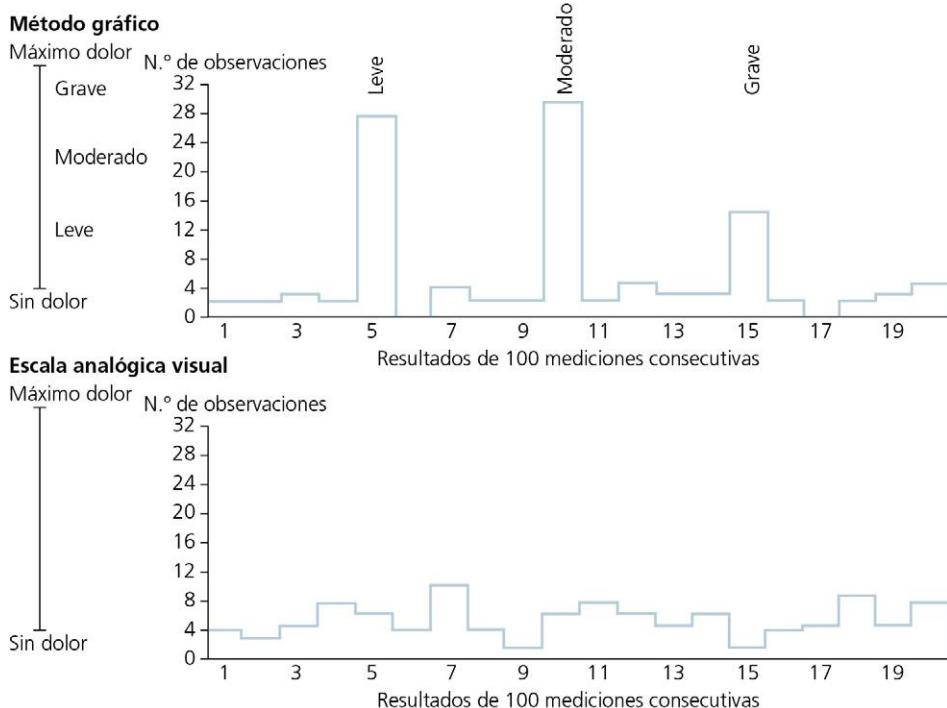


Figura 19.1 Métodos de representación gráfica del dolor (ejemplo 19.3).

#### Cuadro 19.2 Características de una buena escala de medida

- Apropriada para su uso en el estudio, de acuerdo con los objetivos y la definición de la variable
- Viable, de acuerdo con los métodos que podrán utilizarse para recoger la información
- Con suficiente potencia para alcanzar los objetivos del estudio
- Categorías claramente definidas
- Número suficiente de categorías, pero no innecesariamente elevado
- Exhaustiva en su conjunto para permitir clasificar todas las situaciones posibles
- Categorías mutuamente excluyentes (cada situación debe clasificarse en una única categoría)
- Categorías ordenadas jerárquicamente
- Capacidad para medir tanto la mejoría como el empeoramiento del síntoma o la enfermedad en estudio

el mercado. Se obtuvo una concordancia bastante buena (90%) en la duración en meses del consumo, pero bastante menor en cuanto a la duración y la marca (62%) y la dosis (54%).

#### Cuadro 19.3 Fuentes de información

- Observación directa:
  - Exploración física
  - Exploraciones complementarias
- Entrevistas y cuestionarios
- Registro de datos por el propio paciente
- Informador indirecto
- Registros y documentos ya existentes (datos secundarios):
  - Datos individuales
  - Datos agregados

#### Observación directa

La utilidad del examen físico o de las medidas biológicas depende, principalmente, de su estabilidad en el tiempo. En un extremo se encuentran variables como el grupo sanguíneo, que permanecen fijas durante toda la vida, y en el otro, variables como el monóxido de carbono, que indican el consumo de tabaco en las horas que preceden a su medición. En una situación intermedia estarían, por ejemplo, el peso o la presión arterial.

Un problema que se puede presentar con las medidas biológicas es que su valor se modifique por la presencia de una enfermedad.

**Ejemplo 19.5.** Supongamos un estudio de casos y controles en el que se estudiara el riesgo de padecer infarto agudo de miocardio (IAM) asociado a la presencia de hipertrofia ventricular izquierda. En las horas inmediatas después de padecer un IAM se observa un descenso de las cifras de colesterol, por lo que si la valoración de la colesterolemia se hace en función de una medición de las cifras de colesterol a las pocas horas de padecer el IAM, se producirá un error de medición y un sesgo en la estimación de su efecto sobre el riesgo de desarrollar la enfermedad.

Cuando las mediciones corresponden a pruebas de laboratorio o pruebas complementarias, es importante asegurarse de que los valores no dependen del observador, que los aparatos de medida están bien calibrados, que las unidades de medida son las de referencia y que sus valores normales son conocidos y comparables entre los laboratorios participantes en el estudio.

Muchas variables se pueden modificar mediante una intervención preventiva o terapéutica. Esta consideración es importante en los diseños transversales de asociación cruzada, por ejemplo, entre la presión arterial y la edad, en el supuesto de que los individuos hipertensos hayan recibido alguna intervención para reducir sus cifras de presión arterial.

## Entrevistas y cuestionarios

Permiten obtener información tanto sobre exposiciones ocurridas en el pasado como en la actualidad. Son la fuente de información de elección cuando se requieren grandes cantidades de datos, en especial, si se refieren principalmente a exposiciones pasadas y que han tenido un gran impacto en la vida del individuo. Los capítulos 20 y 21 de esta obra se dedican al diseño de cuestionarios y a su validación.

Cada vez es más frecuente que las variables de respuesta tanto de los ensayos clínicos como de los estudios observacionales sean medidas de la calidad de vida relacionada con la salud (CVRS). El desarrollo y el uso creciente de los instrumentos de la CVRS viene dado por la convicción de que las medidas de resultado tradicionales, como la tasa de mortalidad, no son lo suficientemente sensibles para analizar las posibles diferencias entre tratamientos, y que algunos de ellos, si bien aumentan la supervivencia, pueden tener un impacto negativo sobre la calidad de vida del paciente.

## Registro por el propio paciente

Se refiere a un registro detallado y prospectivo llevado a cabo por los propios sujetos del estudio. El detalle (y la duración del registro) con que se recoge la información está definida por el investigador. En la mayoría de las ocasiones, el registro es diario y su duración es de pocos días o semanas.

Al ser registros prospectivos, no dependen de la memoria de los sujetos, y son muy útiles para exposiciones muy frecuentes y que tengan, por ejemplo, poco impacto en la vida de los individuos. Se han usado para medir la actividad física, la actividad sexual, el consumo de alcohol o la dieta. Por otro lado, al registrar las actividades a medida que se realizan, permite conocerlas con detalle, lo que supone una ventaja frente a otras fuentes de información que recogen simplemente la actividad habitual.

La principal limitación es que sólo se pueden registrar variables que se observan durante la realización del estudio. Por tanto, no sirven para medir una exposición en el pasado, a no ser que esté muy correlacionada con la actual. Además, los diarios requieren más tiempo de dedicación por parte de los sujetos incluidos en el estudio, por lo que éstos deben estar motivados para llevar a cabo el registro. Al recoger gran cantidad de datos, tienen el inconveniente añadido de una mayor dificultad de codificación y procesamiento de dichos datos. Por todas estas dificultades, su uso ha sido muy limitado, en general, como criterio de referencia para estudios de validación de cuestionarios u otros métodos de recogida de datos. Las principales fuentes de error de este tipo de registros se resumen en el cuadro 19.4.

## Informador indirecto

La entrevista a personas próximas a los sujetos incluidos en el estudio se usa cuando éstos no son

Cuadro 19.4 **Errores frecuentes en el uso de los registros por el propio paciente**

- El tiempo de cobertura del registro puede no ser suficiente para reflejar la verdadera exposición del individuo
- El registro puede no reflejar las variaciones en la exposición con el tiempo
- El hecho de mantener un registro diario puede afectar (cambiar) el comportamiento de los participantes en relación con la exposición
- Inexactitudes de los participantes al recoger los datos
- Errores en la codificación

capaces de proporcionar la información necesaria. Las causas más frecuentes de esta incapacidad son la defunción del individuo seleccionado, la presencia de enfermedades mentales o la edad.

Son muy útiles en estudios de casos y controles sobre una enfermedad de elevada letalidad y cuando la serie de casos es muy pequeña. En estas circunstancias, la falta de información de los individuos que han muerto puede causar un sesgo importante y, por consiguiente, es útil el estudio de la historia de exposición a través de sujetos próximos a él. Su uso puede aumentar el número de sujetos disponibles y conseguir una muestra más representativa.

Este método de obtención de información presenta limitaciones añadidas a las de las entrevistas personales. Es posible que la persona que responde no sepa con exactitud la historia de exposición del individuo. Además, si la razón por la cual ha de responder es por la muerte de una persona, es posible que este hecho pueda alterar las respuestas. Con el fin de atenuar estos errores, es conveniente que los informadores indirectos sean personas muy próximas al individuo incluido en el estudio.

La fiabilidad de los datos proporcionados por un informador indirecto varía en función de la variable sobre la que se recoge la información. Por ejemplo, la fiabilidad es alta cuando se pregunta sobre los estudios, y moderada si es sobre el consumo de tabaco o la dieta.

En ocasiones, el uso de una persona próxima puede proporcionar información más fiable que el propio sujeto, por ejemplo, si se interroga a una madre sobre la historia de las enfermedades padecidas por su hijo en la infancia.

## Registros previos

Los registros previos contienen datos obtenidos para otro propósito que no es el del estudio de investigación, y se denominan *datos secundarios*, por oposición a los *datos primarios*, recogidos directamente para la realización del estudio. El cuadro 19.5 presenta algunos ejemplos de registros utilizados frecuentemente en investigación.

Sus principales ventajas radican en que son fuentes de datos rápidas, sencillas y económicas. Además, si los registros son exhaustivos, no habrá pérdidas de información debido a las no respuestas. Si la información que contienen se ha registrado prospectivamente, se minimizan errores de memoria.

Tienen importantes limitaciones relacionadas fundamentalmente con su validez y calidad. Los datos que contienen han sido recogidos por múltiples personas, que pueden haber utilizado definiciones y métodos diferentes. Además, aunque los datos sean

Cuadro 19.5 Ejemplos de datos secundarios

- Censo poblacional
- Estadísticas de mortalidad
- Estadísticas de natalidad
- Estadísticas demográficas
- Registros laborales
- Encuestas poblacionales
- Registros de malformaciones
- Estadísticas de centros de control epidemiológico
- Enfermedades de declaración obligatoria
- Altas hospitalarias
- Registros hospitalarios
- Registros de actividad de los centros de salud
- Historias clínicas
- Datos de estudios previos

homogéneos, pueden no corresponder a la definición concreta que el investigador desea utilizar. Por ejemplo, en muchas historias clínicas no se registra el número de cigarrillos que fuma habitualmente una persona o los gramos de alcohol que bebe. Existen problemas añadidos relacionados con la ilegibilidad de algunas caligrafías o la dificultad para encontrar determinada información enmascarada entre otros muchos datos irrelevantes para el estudio.

Las bases de datos se mantienen habitualmente con finalidades clínicas o administrativas, pero no de investigación, por lo que no suelen recogerse con la debida meticulosidad. Una decisión que debe tomarse con la información que falta es si se debe considerar que el individuo no tiene la exposición o la enfermedad, o si se debe considerar una pérdida de información. En enfermedades (o exposiciones) graves y poco frecuentes, se puede asumir que la falta de información es equivalente a la ausencia de la condición. Por ejemplo, en un trabajo en el que se estudian los aneurismas de aorta, si la información no especifica claramente que el individuo tiene esta enfermedad, se puede codificar como que no la tiene. Por el contrario, si se trata de exposiciones frecuentes, como el consumo de tabaco, la falta de información debe ser tratada como tal y no como si el individuo no fumara.

Antes de utilizar una determinada base de datos, deben conocerse las definiciones empleadas, el método de recogida y procesamiento de los datos y su validez.

Pueden diferenciarse dos grandes tipos de datos secundarios:

- *Datos individuales*, que proporcionan información separadamente para cada sujeto. Provienen

- sobre todo de la documentación clínica de los hospitales y los centros de salud, de registros de determinadas enfermedades o procesos, o de datos recogidos en estudios previos.
- **Datos agregados**, que proporcionan información sobre grupos de individuos, pero no de forma separada para cada uno de ellos. Su principal inconveniente radica en la posibilidad de incurrir en una falacia ecológica, ya que las asociaciones observadas en los datos agregados no son necesariamente ciertas individualmente.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- 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 resultados clínicos. *Med Clin (Barc)*. 1995;104:771-6.
- Rosenberg MJ, Layde PM, Ory HW, Strauss LT, Rooks JB, Rubin GL. Agreement between women's histories of oral contraceptive use and physician records. *Int J Epidemiol*. 1983;12: 84-7.
- Thompson C, Briggs M. Support for carers of people with Alzheimer's type dementia. *Cochrane Database Syst Rev*. 2000;(2): CD000454.
- Walter SD, Clarke EA, Hatcher J, Stitt LW. A comparison of physician and patient reports of Pap smear histories. *J Clin Epidemiol*. 1988;41: 401-10.
- Rosenberg MJ, Layde PM, Ory HW, Strauss LT, Rooks

## BIBLIOGRAFÍA

- Armstrong BK, White E, Saracci R. Principles of exposure measurement in epidemiology. Oxford: Oxford University Press; 1994.
- Badia X, Salamero M, Alonso J. La medida de la salud: guía de escalas de medición en español. 3.<sup>a</sup> ed. Barcelona: Edimac; 2002.
- Doyle JR, Doyle JM. Measurement error. Measurement error is that which we have not yet explained. *BMJ*. 1997;314: 147-148.
- Howe GR. Use of computerized record linkage in cohort studies. *Epidemiol Rev*. 1998;20:112-9.
- Hyland ME. A brief guide to the selection of quality of life instrument. *Health Qual Life Outcomes*. 2003;1(1):24.
- Latour J, Abraira V, Caballero JB, López Sánchez J. Las mediciones clínicas en cardiología: validez y errores de medición. *Rev Esp Cardiol*. 1997;50:117-28.
- Michels KB. A renaissance for measurement error. *Int J Epidemiol*. 2001;30:421-2.
- Moran LA, Guyatt GH, Norman GR. Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. *J Clin Epidemiol*. 2001;54:571-9.
- Muñoz A, Gange SJ. Methodological issues for biomarkers and intermediate outcomes in cohort studies. *Epidemiol Rev*. 1998;20:43-56.
- Epidemiol Rev. 1998;20: 29-42.
- Nelson LM, Longstreth Jr WT, Koepsell TD, Van Belle G. Proxy respondents in epidemiologic research. *Epidemiol Rev*. 1990;12:71-86.
- Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology*. 2003;14:459-66.
- Stang A. Appropriate epidemiologic methods as a prerequisite for valid study results. *Eur J Epidemiol*. 2008;23(12):761-5.
- White E, Hunt JR, Casso D. Exposure measurement in cohort studies: the challenges of prospective data collection. *Epidemiol Rev*. 1998;20:43-56.

## Diseño de cuestionarios

Un cuestionario es un proceso estructurado de recogida de información a través de la cumplimentación de una serie predeterminada de preguntas. Estructurado significa que a todos los participantes se les formulan las mismas preguntas, de la misma forma y en la misma secuencia. La finalidad es conseguir la máxima fiabilidad en la información.

Existe una gran cantidad de instrumentos o escalas disponibles, pero es frecuente que los investigadores rechacen los cuestionarios empleados en otros estudios, aduciendo que ninguno de ellos es totalmente adecuado, y desarrollen uno nuevo, con la esperanza de que será más válido y fiable, desestimando los esfuerzos y dificultades que supone elaborar y validar un nuevo cuestionario. Además, esta multiplicidad de escalas dificulta la comparación de resultados entre estudios.

Las etapas que se siguen en la elaboración de un nuevo cuestionario se muestran en el cuadro 20.1.

### TIPOS DE CUESTIONARIOS

La forma de obtener la información a través de un cuestionario puede ser de dos tipos: que sea el propio participante quien lo cumplimente (encuestas autocumplimentadas) o que sea administrado por un encuestador, ya sea personalmente o por teléfono.

#### Encuestas autocumplimentadas

Al no mediar un entrevistador, no se introducen sesgos por la forma de formular las preguntas ni por el modo de registrarlas, y además el estudio es más económico. Otra ventaja es que la duración

Cuadro 20.1 **Etapas en la preparación de un cuestionario**

1. Decidir la información necesaria y revisar la bibliografía
2. Elegir el tipo de cuestionarios
3. Elegir el tipo de preguntas para cada variable
4. Definir códigos, puntuaciones y escalas
5. Elegir el orden de las preguntas
6. Diseñar el formato
7. Preparar el manual de instrucciones
8. Entrenar a los encuestadores
9. Realizar una prueba piloto
10. Revisar el cuestionario y el manual de instrucciones

del estudio será menor, ya que se puede obtener información de un gran número de personas sin necesidad de entrevistarlas personalmente.

Cuando el cuestionario se envía por correo, su principal desventaja es que el porcentaje de respuestas es muy bajo comparado con los otros métodos. Este porcentaje varía según el problema estudiado y la motivación de las personas seleccionadas para la encuesta.

Una estrategia para mejorar el número de respuestas es enviar el cuestionario dos y hasta tres veces a las personas que no lo devuelven y/o efectuar una llamada telefónica invitando a su cumplimentación y devolución. Aun así, es de esperar que el número de personas que respondan sea menor que en las entrevistas personales.

Cuando se estudian personas que pertenecen a grupos muy definidos, como escuelas o centros sanitarios, puede darse el cuestionario personalmente

a cada individuo seleccionado y recogerlo un tiempo después.

Las preguntas han de ser sencillas y las instrucciones muy claras. Si el cuestionario es complejo se corre el riesgo de que muchas preguntas queden sin contestar, o lo que es peor, que se respondan erróneamente por falta de comprensión. Además, serán de poca utilidad cuando la muestra de personas seleccionadas no posea un grado de escolarización suficiente como para leer y escribir las respuestas con fluidez. El cuadro 20.2 recoge las ventajas y desventajas mencionadas.

### **Encuestas administradas por un encuestador**

El entrevistador puede aumentar el porcentaje de respuestas y motivar al participante a responder correctamente, o clarificar y obtener datos más completos. Por otro lado, puede introducir un sesgo al preguntar, ya sea omitiendo cuestiones o cambiando sus palabras, al clarificar las respuestas; ya sea porque no lo hace o lo hace de forma inadecuada o sesgada, al registrar las respuestas dadas por el entrevistado; ya sea porque falsea los datos y registra respuestas de preguntas (o todo un cuestionario) que no se han formulado.

**Cuadro 20.2 Características de los cuestionarios autocumplimentados**

#### **Ventajas**

- Bajo coste en relación con los cuestionarios por entrevista personal
- Mayor posibilidad para encuestar a un gran número de personas
- Se elimina el sesgo debido al entrevistador
- Contestación más cómoda para el encuestado
- Mayor estandarización en la presentación del material

#### **Desventajas**

- Requieren un mínimo grado de escolarización y capacidad de lectura
- La limitación habitual radica en que las preguntas han de ser simples y cerradas
- Incapacidad de clarificar preguntas o respuestas
- Si se envían por correo, hay un alto porcentaje de no respuestas
- No permiten observar respuestas emocionales
- Ausencia de seguridad sobre quién responde y si lo hace solo

**Ejemplo 20.1.** Supongamos que en un estudio para validar los datos sobre el uso de servicios sanitarios en que la información se recoge a través de una entrevista personal, se observa que la frecuencia registrada en el uso de los servicios sanitarios desciende en los cuestionarios cumplimentados a partir de la quinta semana desde que se inició el estudio, lo que podría indicar una pérdida de interés y entusiasmo por parte de los entrevistadores a medida que avanza el estudio. Éste es un ejemplo de un sesgo debido al entrevistador.

La variación debida al entrevistador no puede ser controlada en su totalidad. Incluso si se emplea a un solo encuestador, éste puede variar su técnica de entrevista durante el estudio. Por ello, la selección y la formación de los entrevistadores es un paso clave, que permite estandarizar la manera de formular las preguntas y de registrar las respuestas. Estos aspectos se comentan con detalle en el capítulo dedicado a la gestión del estudio (ver capítulo 24).

### **Entrevistas personales**

Éste es el tipo de encuestas que proporciona el mayor número de respuestas. Es más fácil contestar unas preguntas verbalmente que por escrito. Además, en un cuestionario por correo los participantes pueden dejar sin contestar parte de una sección y pasar a la siguiente. En el cuadro 20.3 se resumen sus características.

Las preguntas pueden ser más complejas y la información más exhaustiva. La presencia de un encuestador facilita la respuesta a personas con

**Cuadro 20.3 Características de los cuestionarios para entrevistas personales**

#### **Ventajas**

- Permiten alcanzar un alto porcentaje de respuestas
- Mayor control sobre la secuencia de las preguntas
- Permiten clarificar las preguntas y las respuestas
- Las preguntas pueden ser más complejas
- El número de preguntas puede ser elevado, con lo que se obtiene una información más exhaustiva

#### **Desventajas**

- Alto coste
- Se necesita más tiempo para terminar el estudio
- Posibilidad de introducir sesgos debidos al entrevistador

dificultades para leer y escribir. Permite, también, reconducir la entrevista en el supuesto de que el encuestado se muestre distraído o confuso.

Un error frecuente es que no todos los encuestadores siguen las mismas técnicas de entrevista, con lo que se pierde uniformidad en el proceso de recogida de datos. La actitud del entrevistador es percibida por los sujetos; las palabras que usa al formular una pregunta y la entonación que les da son factores que influyen sobre las respuestas. Los participantes, a menudo, desean complacer a los encuestadores y dan una respuesta que piensan que merecerá su aprobación.

### Encuestas telefónicas

El uso de la encuesta telefónica ha ido en aumento en los últimos años. Muchas de las ventajas de las entrevistas personales también se consiguen con las telefónicas. Tienen un coste intermedio y el porcentaje de respuestas es más alto que en las encuestas por correo, aunque algo más bajo que en las realizadas por entrevista personal.

La entrevista ha de ser breve y muy sencilla. Estar muchos minutos al teléfono respondiendo preguntas que por su complejidad requieren frecuentes aclaraciones es cansado y hace que el entrevistado termine colgando el auricular. Hay que evitar las preguntas íntimas porque muy probablemente la gente se muestre reacia a responderlas por teléfono.

Con el fin de conseguir un buen número de respuestas es indispensable, al igual que en cualquier otro tipo de encuesta, enviar una carta previa donde se expliquen los motivos por los que se realiza el estudio y la institución que lo efectúa. Es primordial, además, la presentación telefónica. Hay que decir el nombre del encuestador y repetir los fines del estudio, el nombre de la institución que lo respalda y cómo son seleccionados los participantes.

El éxito de la encuesta dependerá del número de personas candidatas que no tengan teléfono, y si

difieren sustancialmente respecto a las características que se van a estudiar. Si el número es muy alto y las diferencias importantes, se obtendrá un resultado sesgado.

## TIPOS DE PREGUNTAS

Existen dos tipos de preguntas: las cerradas y las abiertas ([tabla 20.1](#)).

### Preguntas cerradas

Las preguntas cerradas son aquellas en que se especifica de antemano las posibles respuestas alternativas. Son más fáciles de aplicar y, también, de tabular y analizar que la mayoría de las preguntas abiertas. La recogida de datos gana en precisión y uniformidad, a la vez que es más eficiente, ya que una persona puede contestar, por norma general, un mayor número de preguntas cerradas que abiertas, en un tiempo determinado. Además, si se leen las posibles respuestas, pueden ayudar al entrevistado a recordar o a considerar opciones que de otra forma le hubieran pasado inadvertidas.

Las preguntas cerradas son más difíciles de elaborar, ya que hay que considerar todas las posibles opciones alternativas en términos correctos y comprensibles para todos los participantes. Otra desventaja es que permiten sólo un número limitado de opciones, sin que los encuestados puedan matizarlas o proporcionar información adicional.

### Preguntas abiertas

En las preguntas abiertas las respuestas no están preestablecidas. Las respuestas suelen ser más detalladas, ya que se recoge íntegramente lo que los participantes expresan libremente con sus propias palabras. Sin embargo, es difícil medir diferencias

**Tabla 20.1 Ventajas y desventajas de las preguntas abiertas y cerradas**

Preguntas cerradas	Preguntas abiertas
<b>Ventajas</b> <ul style="list-style-type: none"><li>- Obligan a reflexionar sobre detalles</li><li>- Uniformidad de respuesta</li><li>- Facilidad de codificación</li></ul>	<ul style="list-style-type: none"><li>- Útiles para informaciones complejas</li><li>- Permiten observaciones y comentarios</li><li>- No se inducen respuestas</li></ul>
<b>Desventajas</b> <ul style="list-style-type: none"><li>- No son útiles para informaciones complejas</li><li>- Inducción de las respuestas</li><li>- Pueden no recoger datos importantes</li></ul>	<ul style="list-style-type: none"><li>- Requieren mayor tiempo y esfuerzo</li><li>- Dificultad para codificar</li><li>- No son útiles cuando se pregunta sobre actitudes o hechos mal vistos socialmente</li></ul>

entre los participantes, ya que una misma pregunta puede ser contestada con diferentes frases, lo que dificulta su interpretación. Las preguntas abiertas son más útiles cuando la información que se desea recoger es muy compleja, de forma que si se quisieran utilizar preguntas cerradas ello obligaría a predefinir multitud de opciones, o cuando se usan con finalidades exploratorias porque se desconocen cuáles serían las opciones de respuesta más adecuadas.

El análisis de las preguntas abiertas suele ser lento, y si se decide tabular las respuestas, puede ser difícil decidir en qué categoría hay que clasificar una determinada respuesta. Es posible que el investigador decida más en función de sus opiniones o preferencias que en las de los propios participantes.

Existen preguntas abiertas en las que la respuesta es directa, como la edad, el lugar de nacimiento o de residencia, el número de cigarrillos fumados por día, etc. El uso de preguntas cerradas para este tipo de cuestiones conlleva una pérdida de información y, para conductas mal consideradas socialmente, un mayor grado de error.

**Ejemplo 20.2.** En un estudio en el que se formulaba una pregunta cerrada sobre el número de horas al día en que se miraba la televisión (Schwartz et al., 1985), un 16% estimó que la miraba más de 2 horas y media. Cuando se aumentó el número de las categorías, un 37,5% estimó que la miraba más de 2 horas y media. En general, muchos participantes huyen de los valores extremos y escogen los intermedios, ya que piensan que esto es lo que contestará la mayoría. Si la pregunta hubiera sido abierta se hubiera soslayado este problema.

## SELECCIÓN DE LOS ÍTEMES

El diseño de un cuestionario generalmente se inicia con la selección de los ítems que luego se traducirán a preguntas. Hay que recoger la cantidad mínima de datos necesarios para alcanzar los objetivos del estudio.

Entre los aspectos que limitan la viabilidad de la administración de un cuestionario se encuentra su duración. Se estima que un cuestionario mediante entrevista no debe durar más de 60-90 minutos, y por teléfono, no más de 40 minutos. Si un cuestionario enviado por correo es muy largo, es posible que las personas que deban responderlo piensen que será demasiado pesado y decidan no rellenarlo. Por esta razón se recomienda que los cuestionarios autocumplimentados no sobrepasen las 12 páginas (Dillman, 1978).

**Ejemplo 20.3.** En un estudio se comparaba la administración de una versión completa de un cuestionario con otra abreviada a un grupo de estudiantes (Herzog y Batchman, 1981). Las personas que respondieron la versión completa tenían más tendencia a dar las mismas opciones de respuesta a la mayoría de los ítems (p. ej., contestaban siempre de acuerdo o en desacuerdo) que aquellos a los que se les administró la versión abreviada.

Otros aspectos que dificultan las respuestas son el período de tiempo al que se refieren las preguntas, el impacto que ha podido tener en la vida del individuo el hecho por el que se le pregunta, y la complejidad y el detalle de los datos que se solicitan. Si los encuestados tienen que recordar hechos que han sucedido muchos años atrás, y/o se trata de hechos que han tenido poco impacto y/o son muy complejos de explicar, muchos tenderán a no contestar el cuestionario.

Como consecuencia de los aspectos mencionados en los párrafos anteriores, existe el riesgo de que los participantes den por terminada la entrevista antes de tiempo o, si se trata de un cuestionario autocumplimentado, de que dejen una parte sin contestar. Además, existe la posibilidad de que los datos obtenidos sean de poca calidad, de que disminuya el porcentaje de respuestas y de descrédito de las encuestas y, en consecuencia, que la población tienda a no aceptar participar en ellas. Este último problema es especialmente importante en los estudios longitudinales que requieren encuestas seriadas durante el proceso.

La decisión sobre si las preguntas deben tener o no *validez lógica*, es decir, si su significado y relevancia es evidente, debe tomarse antes de iniciar su redacción. Si carecen de validez lógica es muy probable que los encuestados rechacen contestar. De todos modos, en alguna ocasión puede ser de interés disimular la verdadera finalidad de los ítems por temor a que el encuestado trate de amañar las respuestas. Por ejemplo, es probable que un paciente trate de aparentar que está más enfermo de lo que realmente está, con el fin de conseguir una mejor o más rápida asistencia médica.

## REDACCIÓN DE LAS PREGUNTAS

Decidir cómo se van a formular las preguntas es una de las tareas más difíciles. Una forma apropiada de empezar es escribir una lista con los ítems seleccionados (información que se desea obtener). A partir de aquí irán surgiendo las distintas preguntas que abarquen cada uno de estos conceptos. Es frecuente que el número de preguntas se vaya ampliando a medida que avanza el diseño del cuestionario.

Por ello, ya en las fases iniciales, es muy útil definir el plan de análisis con el fin de imaginar los resultados que se obtendrán en función de su posible utilidad, y decidir si puede suprimirse alguna pregunta.

Hay que prestar mucha atención a la redacción de las preguntas, muy especialmente cuando se trata de actitudes (predisposición a hacer algo), comportamientos (conductas) y creencias (convicciones en relación con el sistema de valores del individuo). El entrevistado puede entremezclar lo que piensa, lo que haría y lo que realmente hace. En este sentido, la sola modificación del tiempo de un verbo puede cambiar el significado de una pregunta, y lo que era un comportamiento pasa a ser una actitud o una creencia. En el cuadro 20.4 se resumen algunas recomendaciones prácticas.

### Evitar preguntas ambiguas

Una pregunta no ha de admitir más de una interpretación. Incluso las preguntas que parecen más sencillas y directas pueden ser ambiguas para quienes responden la encuesta.

**Ejemplo 20.4.** Supongamos que a un hipertenso se le pregunta: ¿qué tratamiento sigue usted? Ante esta pregunta es difícil que sepa si debe contestar el tratamiento que le recomendó su médico, con independencia de su cumplimiento, o bien el trata-

Cuadro 20.4 Aspectos prácticos a tener en cuenta en la redacción de las preguntas

- Evitar preguntas ambiguas
- No usar términos vagos como: en ocasiones, a menudo, etc.
- No formular las preguntas en forma negativa
- No formular dos preguntas en una
- Las preguntas deben ser cortas y no contener muchos conceptos (un concepto, una pregunta)
- Utilizar un lenguaje sencillo
- No sobrevalorar la memoria de los encuestados
- Utilizar preguntas cerradas para las cuestiones personales
- Las preguntas han de ser neutras
- Evitar un número de opciones elevado en las preguntas cerradas
- Considerar todas las opciones posibles en las preguntas cerradas
- En las preguntas cerradas, cuando el encuestado sólo debe elegir una de las opciones, éstas deben ser mutuamente excluyentes
- Ordenar lógicamente las alternativas

miento que hace en realidad, independientemente del que le recomendó su médico.

### No usar términos vagos

Se debe evitar el uso de términos vagos o ambiguos como *en ocasiones, a menudo, etc.*

**Ejemplo 20.5.** La pregunta *¿consulta a menudo a su médico por dolor de cabeza?* sería mejor sustituirla por otra más precisa, como *¿cuántas veces consultó a su médico por dolor de cabeza durante las dos últimas semanas?*

### No formular las preguntas de forma negativa

Las preguntas negativas conducen a dobles interpretaciones. Considérese la siguiente pregunta: *¿no piensa usted que fumar perjudica su salud?* Una respuesta afirmativa puede significar: sí, creo que perjudica mi salud; o bien: sí, no pienso que perjudique mi salud.

### No formular dos preguntas en una

Considérese la siguiente pregunta: *¿ha tenido alguna vez dolor en el pecho cuando sube escaleras o una cuesta?* La respuesta puede ser negativa sencillamente porque el encuestado no acostumbra a subir escaleras. Si se quiere conocer si una persona ha tenido dolor en el pecho, si este dolor se produce cuando sube escaleras, o si se calma con el reposo, es conveniente formular una pregunta para cada uno de estos conceptos. Se empezaría preguntando: *¿ha tenido alguna vez dolor en el pecho?* Esta cuestión actuaría como filtro, y si la respuesta fuese negativa, no haría falta efectuar las restantes preguntas. El uso de cuestiones filtro conlleva el desarrollo de instrucciones complejas. Esta dificultad no es importante cuando los entrevistadores están bien entrenados y adquieren experiencia en el manejo del cuestionario.

**Ejemplo 20.6.** En un cuestionario sobre estilos de vida, se incluían las siguientes preguntas sobre el hábito de fumar:

Pregunta 33. *¿Es usted fumador/a de cigarrillos, puros o pipa?*

Sí ( ) No ( )

En caso de respuesta negativa pase a la pregunta número 44.

Pregunta 34. *¿Actualmente fuma?*

Sí ( ) No ( )

En caso de respuesta negativa pase a la pregunta 39.

## Un concepto, una pregunta

Las preguntas no han de contener muchos conceptos. Considérense las siguientes preguntas: ¿ha notado alguna vez los tobillos o los pies hinchados al levantarse por la mañana, o a últimas horas del día?, ¿ha notado alguna vez los tobillos o los pies hinchados? La segunda sería preferible, ya que sólo expresa una idea, por lo que no se distrae la atención del encuestado.

## Utilizar un lenguaje sencillo

El lenguaje debe ser comprensible para todos los participantes del estudio, incluidos aquellos con un menor grado de escolarización, y no hay que usar terminología médica. Tampoco deben utilizarse abreviaturas, siglas, extranjerismos, argots, etc. Una forma de solucionar este problema es incluir notas aclaratorias con sinónimos más coloquiales entre paréntesis.

**Ejemplo 20.7.** En un cuestionario interesa recoger la fecha de la menarquia. Una posible pregunta es: ¿a qué edad tuvo la menarquia (la primera menstruación)?

La regla general es que la escala no debería exigir una habilidad de lectura más allá de la que tiene una persona de 12 años. Las frases deben ser cortas y con una estructura gramatical lo más sencilla posible.

## No sobrevalorar la memoria de los encuestados

El recuerdo varía según la naturaleza del problema y el tiempo transcurrido. Estancias breves en el hospital a causa de una enfermedad crónica, o medicaciones que se toman diariamente, son hechos importantes que pueden olvidarse sólo porque, para el encuestado, no son nuevos. En este caso, el uso de una lista de enfermedades o de medicaciones será de gran ayuda para estimular la memoria del encuestado.

**Ejemplo 20.8.** Considérese la siguiente pregunta: ¿cuántas veces ha consultado a su médico en el último año? Éste es un dato que mucha gente no recordará. Sin embargo, si se pregunta por un período más reciente, por ejemplo, las dos últimas semanas, es más fácil de recordar.

## Utilizar preguntas cerradas para las cuestiones personales

Las preguntas personales que pueden inculpar al sujeto o ir en contra de los valores de ciertos sectores de la sociedad han de redactarse con sumo cuidado, ya que tienden a conducir a un rechazo o a que se

responda erróneamente. Es mejor utilizar preguntas cerradas entre cuyas respuestas alternativas se enmascare la información que se deseé obtener.

**Ejemplo 20.9.** Para un encuestado puede ser muy embarazoso contestar directamente a la pregunta de si ha padecido una determinada enfermedad venérea en los últimos años. Sin embargo, si esta enfermedad se encuentra como una opción dentro de una lista de enfermedades, se facilita la respuesta:

¿El año pasado padeció usted alguna de las siguientes enfermedades?

	Sí	No	NS/NC
Infección de orina	( )	( )	( )
Cálculos en los riñones	( )	( )	( )
Gonorrea	( )	( )	( )

## Las preguntas han de ser neutras

Deben evitarse preguntas que sugieran algún tipo particular de respuesta. El uso de palabras con un alto contenido emocional o ideológico (p. ej., libertad, democracia, saludable, etc.) condiciona respuestas positivas, mientras que otros términos (burocracia, inspección, control, etc.) pueden llevar a respuestas negativas. Del mismo modo, el uso de atributos o adjetivos en el enunciado puede influir en la respuesta.

**Ejemplo 20.10.** Si a alguien se le pide que dé su opinión sobre la siguiente frase: «desde el punto de vista sanitario es indudable que se debería prohibir fumar en los restaurantes», la palabra indudable puede inducir a los encuestados a mostrar un total acuerdo con esta proposición.

Hay que ser muy cuidadoso cuando se pregunta sobre conductas o actitudes que no están bien aceptadas socialmente. Ante estas preguntas, el participante es más proclive a falsear las respuestas. En estas circunstancias, cualquier palabra, o la entonación del entrevistador, puede ser motivo para que el encuestado no exprese su propia opinión, y conteste lo que «todo el mundo espera que conteste».

## Evitar un número de opciones muy elevado en las preguntas cerradas

El número de respuestas alternativas debería ser menor de diez. En caso contrario, se corre el peligro de que el encuestado no las recuerde todas y tienda

a escoger una de las últimas. Sin embargo, cuando se pregunta si ha padecido alguna enfermedad o si toma algún medicamento, el número de opciones suele ser elevado. En estas ocasiones, es útil que el encuestador muestre una tarjeta en la que se reproduzcan las distintas alternativas con el fin de facilitar la respuesta, ya que el encuestado podrá leerlas.

Además del número, es conveniente que las respuestas alternativas no sean muy largas porque es molesto para los participantes tener que leer tantas palabras o tratar de recordar tantos detalles.

### Considerar todas las posibles alternativas

El investigador, en ocasiones, descuida o no considera algunas opciones potencialmente importantes.

**Ejemplo 20.11.** Consideremos la siguiente pregunta: ¿dónde tuvo lugar la última consulta que hizo usted al médico?

- ( ) Centro de salud.
- ( ) Consulta externa de un hospital.
- ( ) Servicio de urgencias de un ambulatorio.
- ( ) Servicio de urgencias de un hospital.
- ( ) Consulta privada de un médico.

Aunque a primera vista pueda parecer una lista completa, muchas personas pueden no entrar en ninguna de estas categorías, porque han consultado al médico de su empresa o por teléfono, por citar dos ejemplos.

Para evitar estos problemas se suele incluir una categoría denominada *otros*, y se solicita al encuestado que especifique la respuesta. Durante la realización de la prueba piloto pueden detectarse estas omisiones. Si se observa que muchos encuestados eligen la opción *otros*, hay que revisar y completar las opciones de respuesta.

### Las alternativas han de ser mutuamente excluyentes

Si el encuestado sólo puede elegir una opción, las alternativas deben ser mutuamente excluyentes.

**Ejemplo 20.12.** Consideremos la siguiente pregunta: ¿cuántas mamografías le han hecho en los últimos 3 años?

- ( ) Ninguna.
- ( ) 1 o 2.
- ( ) 2 o 3.
- ( ) Más de 3.

Ante esta pregunta, una mujer a la que se le han practicado dos mamografías no sabría si debe contestar la segunda o la tercera opción. Aunque en este ejemplo el error parece evidente, no es raro descubrir este tipo de errores cuando se realiza la prueba piloto.

### Ordenar lógicamente las alternativas

Las distintas alternativas deben seguir un orden lógico, sobre todo cuando las opciones siguen una línea de continuidad o jerarquía.

**Ejemplo 20.13.** ¿Cómo valora la forma en que el médico atiende a las explicaciones que usted le da?

- ( ) Muy satisfactoria.
- ( ) Bastante satisfactoria.
- ( ) Aceptable.
- ( ) Bastante insatisfactoria.
- ( ) Muy insatisfactoria.

## PUNTUACIONES Y ESCALAS

Los códigos y las puntuaciones transforman las respuestas en variables que pueden ser tabuladas y analizadas estadísticamente.

Las respuestas dicotómicas, dada su simplicidad, se emplean con frecuencia en cuestionarios muy largos. A menudo incluyen una tercera categoría: no sabe/no contesta. Sus principales limitaciones son la pérdida de información y el hecho de que algún participante no quiera contestar alguna de las preguntas, ya que no puede matizar la respuesta. Las preguntas cerradas con múltiples opciones tratan de superar estos inconvenientes. Las variables también se pueden medir en una escala ordinal o cuantitativa.

Para poder medir algunos fenómenos se requiere efectuar varias preguntas interrelacionadas. Desde un punto de vista estadístico se pueden usar escalas más complejas para combinar las diferentes respuestas en una única puntuación. Para conseguir esta puntuación global, todas las preguntas deben evaluar la misma característica. El tipo de escala más usado es la de Likert, aunque también se emplean la de Guttman y la escala visual analógica, entre otras.

### Escala de Likert

Esta escala comprende varias frases y los encuestados deben expresar su grado de acuerdo o desacuerdo con cada una de ellas.

**Ejemplo 20.14.** A continuación encontrará una serie de afirmaciones. Indique el grado de acuerdo o

	Total acuerdo	Acuerdo	No estoy seguro	Desacuerdo	Total desacuerdo
La publicidad del tabaco debería prohibirse	1	2	3	4	5
Hay demasiadas restricciones para fumar	1	2	3	4	5
Hoy en día se insiste demasiado sobre el daño del tabaco	1	2	3	4	5

desacuerdo con cada una de ellas. Por favor, marque con un círculo la opción elegida.

El primer paso consiste en reunir un buen número de frases, entre 10 y 20, que señalen actitudes favorables y desfavorables sobre un tema. Habitualmente cada frase tiene cinco alternativas, que van desde el total acuerdo hasta el total desacuerdo. El número de frases que contengan actitudes positivas o negativas ha de ser similar. Es importante no incluir actitudes neutras o muy extremas que generen una aceptación o rechazo definitivos.

A continuación se concede una puntuación a las respuestas dadas en cada pregunta. En general, la puntuación más alta es para la aceptación de las actitudes consideradas como positivas, o el desacuerdo con las consideradas negativas. La suma aritmética de las puntuaciones servirá para clasificar a los participantes y permitir comparaciones entre ellos. La suma de todas las puntuaciones proporciona una puntuación ordinal total que no garantiza la linealidad o igualdad de los intervalos.

En algunos pocos casos se obtienen «pesos» para cada ítem que reflejan su importancia relativa en el contexto de la escala. En la mayoría de los instrumentos, cada ítem se puntúa igual, ya que se ha demostrado que el uso de «pesos» no mejora la capacidad discriminante o la sensibilidad al cambio del instrumento y, por contra, complica la tarea de calcular las puntuaciones.

### Escala de Guttman

En esta escala se escoge un número relativamente pequeño de frases que expresan una actitud favorable o desfavorable hacia un tema y que están redactadas de tal manera que quien está de acuerdo con una de ellas también lo está con las anteriores.

**Ejemplo 20.15.** Por favor, marque con un círculo cada una de las frases con la que está de acuerdo.

© Elsevier. Fotocopiar sin autorización es un delito.

1. El tabaco puede causar enfermedad.

2. El tabaco es una causa importante de enfermedad.
3. El tabaco es una causa muy importante de enfermedad y muerte.
4. El tabaco es la causa más importante de enfermedad y muerte en España.

En la escala de Guttman las frases son siempre de intensidad creciente y referidas a un solo tema o concepto, y su número no es superior a cuatro o cinco. El hecho de que el número de frases referidas a un concepto sea pequeño hace que este tipo de escalas no discrimine lo suficiente entre individuos y limita su uso generalizado.

### Escala visual analógica

Se presenta a los encuestados una escala en forma lineal con los límites claramente definidos, por ejemplo, entre 0 y 1 o entre 0 y 100. Estos límites son fijos en el sentido de que un extremo representa el estado menos preferido, por ejemplo, el peor estado de salud o más dolor, y el otro el más preferido. Si la línea está dividida en milímetros, se denomina *termómetro*, y si sólo tiene definidos los límites inicial y final, se denomina *escala visual analógica*.

### ORDEN DE LAS PREGUNTAS

Si la secuencia de preguntas sigue un orden lógico se facilitará el proceso de respuesta. Si las preguntas son dispersas y no tienen ninguna ligazón unas con otras se desviarán la atención del encuestado, que irá perdiendo interés a medida que avance la encuesta.

Las primeras preguntas han de despertar el interés del encuestado y aumentar su motivación para colaborar. Las cuestiones más generales, como el estado civil, el trabajo, la dirección o el teléfono, no son buenas para conseguir este propósito y deben dejarse para el final de la encuesta. Igualmente, no hay que iniciar un cuestionario con preguntas de índole

personal o que puedan herir la susceptibilidad de los participantes. Si un encuestado se molesta por una pregunta en concreto, puede dar por terminada la entrevista, y si esta pregunta está al inicio del cuestionario, se perderá toda la información.

La primera pregunta ha de ser sencilla, de tal forma que todos los participantes puedan responderla en pocos segundos; se han de evitar las preguntas abiertas o las cerradas con muchas alternativas. También ha de ser neutral. Un cuestionario no debe empezar con una pregunta en la que los encuestados deban expresar su acuerdo o desacuerdo sobre una cuestión.

Si en un cuestionario se incluyen preguntas abiertas y cerradas sobre un mismo tema, es preferible colocar en primer lugar las abiertas, ya que de esta forma las personas pueden expresar sus opiniones sin estar influidas por las distintas alternativas de las preguntas cerradas.

Toda encuesta debe iniciarse con los planteamientos más generales para pasar posteriormente a los más específicos.

**Ejemplo 20.16.** En un cuestionario sobre la detección precoz del cáncer de mama se incluyan las siguientes preguntas:

- ¿Le han hecho alguna vez una mamografía?
- ¿Por qué se hizo la mamografía?
- El resultado de la mamografía fue...

En esta secuencia de preguntas, la primera actuaría como filtro, ya que si la encuestada respondiera negativamente, no sería necesario contestar las siguientes.

El orden de las preguntas puede influir en las respuestas. Por ello, si se desea comparar los resultados obtenidos al administrar la encuesta en diferentes ocasiones, el orden de las preguntas debe mantenerse constante. Si se desean añadir nuevas preguntas, es preferible colocarlas al final del cuestionario en la medida de lo posible.

Para recoger información sobre sucesos cronológicos (historia laboral, lugares de residencia, etc.), es aconsejable comenzar por el último y retroceder hasta el primero.

El cuadro 20.5 resume los puntos principales que se deben considerar cuando se decide el orden de las preguntas.

## FORMATO DEL CUESTIONARIO

El formato se ha de diseñar prestando atención a los pequeños detalles. Un buen formato evita errores.

Cuadro 20.5 **Guías para decidir el orden de las preguntas**

- Las preguntas han de ser lo suficientemente importantes como para despertar el interés
- Se evitará que la primera pregunta sea una pregunta abierta o una cerrada con muchas alternativas
- La primera pregunta será «neutral»
- Las preguntas sociodemográficas (edad, sexo, grado de estudios, etc.) suelen ir al final del cuestionario
- Las preguntas relacionadas con un mismo tema han de ir consecutivamente
- Si se incluyen preguntas abiertas y cerradas sobre un mismo tema, es preferible colocar en primer lugar las abiertas
- Las preguntas que pueden causar mayor objeción entre los participantes deben ir al final
- Empezar por los planteamientos generales para poder pasar después a los más específicos

Los cuestionarios deben iniciarse con una breve introducción sobre la naturaleza y la finalidad del estudio. En las encuestas personales o telefónicas, el encuestador es quien lee esta introducción. En los cuestionarios autocumplimentados, estará aparte del cuestionario para que el encuestado pueda hojearla cuando desee. Debe incluir la finalidad del estudio, la importancia de que los participantes dediquen unos minutos a responder la encuesta, el nombre de la institución que avala el proyecto, el método de selección de los participantes, el uso que se hará de la información y si se conservará el anonimato y la confidencialidad de los datos, y debe agradecer de antemano la colaboración de los participantes. Al inicio del cuestionario autocumplimentado se darán las instrucciones básicas que se deben seguir, acompañadas de los consiguientes ejemplos.

Es importante no acumular demasiadas preguntas en un espacio reducido. Es preferible que cada una de las opciones esté en una línea.

Es conveniente utilizar frases introductorias cuando se cambia de tema (p. ej., la alimentación es una parte importante en nuestra vida cotidiana; a continuación me gustaría preguntarle acerca de los alimentos que consume habitualmente) o cuando una sección va dirigida a un subgrupo de personas (p. ej., esta sección trata sobre aspectos de salud en la mujer y sólo ha de ser contestada por mujeres). Es conveniente que estos subtítulos queden bien resaltados. Estas anotaciones sirven, además, para romper la monotonía y evitan que el cuestionario parezca un listado interminable de preguntas.

En las preguntas filtro se han de especificar con claridad las instrucciones de cuál es la siguiente pregunta que se debe responder. Ha de quedar claro quién debe contestarla y qué es lo que se supone que debe hacer el que no la ha de responder.

Los números que identifican cada una de las opciones de respuesta se sitúan a la izquierda. La razón principal de ello es que algunas respuestas son más largas que otras, y si el número está a la derecha aumenta la probabilidad de error.

El formato del cuestionario ha de permitir una rápida codificación. La opción más simple es que cada alternativa esté numerada y en el margen derecho del cuestionario se anote el código de la opción seleccionada y se repita el número de la pregunta, con el fin de facilitar la introducción de los resultados en el ordenador.

Las preguntas no se han de partir nunca. Terminar una página con una pregunta a medias crea confusión e invita al error.

En la primera hoja del cuestionario se ha de reservar un espacio para el código de los encuestados. Esto será muy práctico cuando se tenga que verificar y consultar alguna pregunta de un participante en concreto. También suele ser útil anotar el tiempo de duración de la encuesta, así como reservar un espacio para que el encuestador pueda anotar sus propios comentarios.

Una buena pregunta presentada en un buen formato conduce a un fácil registro. El cuadro 20.6 resume algunas recomendaciones. El mejor consejo para conseguir un buen formato es realizar una prueba piloto rigurosa con el fin de detectar los problemas y corregirlos.

## PRUEBA PILOTO

Una vez que se ha completado la primera versión del cuestionario y ya se ha evaluado su legibilidad, éste se administra a un grupo de sujetos representativo de la población a la que va dirigida el instrumento. El tamaño de la muestra de esta prueba piloto debe ser siempre superior al número de ítems, y se recomienda que el número de sujetos se sitúe entre dos y diez veces el número de ítems que tiene el instrumento.

La aplicación del instrumento a estos individuos permitirá obtener los primeros resultados. Se observará, en primer lugar, la frecuencia de respuesta de las distintas opciones de cada ítem. En los cuestionarios sobre conocimientos, esta frecuencia refleja la dificultad de la pregunta. En los cuestionarios sobre opiniones o actitudes, la frecuencia refleja la popularidad de cada ítem. Cuando una alternativa se responde con mucha o poca frecuencia (por en-

### Cuadro 20.6 Recomendaciones para la elaboración del formato de un cuestionario

- Utilizar distintos formatos de letra para cada parte de la pregunta. Por ejemplo, usar:
  - MAYÚSCULAS para la pregunta
  - negrita para las opciones de respuesta
  - cursiva para las instrucciones
- Incluir junto a cada pregunta sus instrucciones específicas
- Usar un formato vertical para las respuestas
- Precodificar todas las preguntas cerradas
- Asociar siempre el mismo código para una misma opción de respuesta (p. ej., utilizar siempre: sí=1)
- Registrar la respuesta a una pregunta cerrada poniendo un círculo o una cruz (pero siempre lo mismo) en la opción escogida
- Proporcionar los espacios necesarios para las preguntas abiertas de respuesta directa
- Recordar que el diseño del formato ha de facilitar la entrada de datos

cima del 90% o por debajo del 10%), discrimina poco entre individuos y, en definitiva, aporta poca información. Este tipo de preguntas no mejora las propiedades psicométricas del cuestionario y además hace que sea más largo, por lo que son candidatas a eliminarse.

**Ejemplo 20.17.** Argimon-Pallàs et al (2010) validaron el test de Fresno, que mide los conocimientos y habilidades de los profesionales sanitarios en medicina basada en la evidencia. Analizaron, entre otras características, la dificultad de las preguntas del test. Para ello decidieron establecer un punto de corte para cada una de las preguntas, por encima del cual se consideraba que el participante «aprobaba» la pregunta. Por ejemplo, en las preguntas cuya puntuación oscilaba entre 0 y 24 puntos, decidieron que una puntuación superior a 19 era un «aprobado». La dificultad de los ítems osciló entre el 15% (sólo un 15% aprobó aquella pregunta) y un 44%. El valor óptimo de dificultad depende del tipo de pregunta, pero si está por debajo del 10 o del 90%, esta pregunta debería revisarse o reemplazarse.

Otro índice de utilidad de un ítem es su *capacidad de discriminación*. Indica si una persona con puntuación elevada en el total del cuestionario es más probable que haya obtenido también una puntuación elevada en un determinado ítem. En otras pala-

bras, se trata de saber si un ítem será de ayuda para discriminar entre los individuos que obtienen una puntuación total alta y los que la obtienen baja.

**Ejemplo 20.18.** Siguiendo con el estudio del ejemplo anterior (Argimon-Pallàs et al, 2010), los autores ordenaron a los participantes según la puntuación total que obtuvieron en el total del test y seleccionaron a los del primer y el último cuartil. Para cada pregunta calcularon la diferencia en el porcentaje de respuestas correctas entre los participantes entre el primer y el último cuartil. Esta diferencia es una medida de discriminación de la pregunta, y en el estudio osciló entre el 7 y el 47%. Una diferencia del 100% hubiera ocurrido si todos los participantes del primer cuartil hubieran contestado correctamente a la pregunta y no lo hubiera hecho ninguno del último cuartil. Se produciría una diferencia de cero cuando el porcentaje de respuestas correctas fuera el mismo en ambos grupos, y esto significaría que la pregunta no discrimina entre grupos.

Otra característica importante es la *homogeneidad* de los ítems, es decir, que midan diferentes aspectos de un mismo fenómeno y no diferentes aspectos de diferentes fenómenos. Esto implica que cada

ítem debe estar moderadamente relacionado con los demás, y a la vez con la puntuación total. Estas dos condiciones son el fundamento de las pruebas de consistencia interna de un cuestionario. Por un lado, si un ítem no está relacionado con otro, quiere decir que probablemente miden fenómenos diferentes y, por tanto, no es lógico sumar sus puntuaciones en una sola escala. Si, por el contrario, dos ítems están muy relacionados quiere decir que la información proporcionada por uno de ellos es redundante.

Una alta consistencia interna de los ítems tiene sentido cuando el interés principal del instrumento reside en describir, por ejemplo, un concepto, un comportamiento o un problema de salud. Por el contrario, cuando el objetivo principal de la escala es discriminar las personas que tienen un determinado atributo de las que no lo tienen, la escala no tiene que ser necesariamente homogénea. Por ejemplo, en un cuestionario de conocimientos sobre una determinada enfermedad, destinado a discriminar a los que saben sobre ella, para alcanzar un rango variado de respuestas correctas entre todos los participantes se tendrán que incluir tanto preguntas fáciles como difíciles, lo que hará disminuir la consistencia interna del cuestionario.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- |   |   |   |
|---|---|---|
| Herzog AR, Batchman JG.<br>Effect of questionnaire<br>length on response quality.<br>Public Opinion Quarterly.<br>1981;45:549-59. | effects of category range<br>on reported behaviour and<br>comparative judgements.<br>Public Opinion Quarterly.<br>1985;49:388-95. | Psychometric properties of a<br>test in evidence based practice:<br>the Spanish version of the<br>Fresno test. BMC Med Educ.<br>2010;10:45. |
| Schwartz N, Hippler HJ, Deutsch B,<br>Strack F. Response scales:  | Argimon-Pallàs JM, Flores-Mateo G,<br>Jiménez-Villa J, Pujol-Ribera E.  |   |

## BIBLIOGRAFÍA

- |  |  |  |
|--|--|--|
| Badia X, Salamero M, Alonso J.<br>La medida de la salud: guía de<br>escalas de medición en español.<br>3. <sup>a</sup> ed. Barcelona: Edimac; 2002.              | Casas Anguita J, Repullo I,<br>Labrador JR, Pereira Candel J.<br>Medidas de calidad de<br>vida relacionadas con la<br>salud. Conceptos básicos,<br>construcción y adaptación<br>cultural. Med Clin (Barc).<br>2001;116:789-96. | J Clin Pract. 2000;54(6):356-9.<br>Dillman D. Mail and telephone<br>surveys. New York: John Wiley<br>& Sons; 1978.   |
| Bosch JLC, Torrente D. Encuestas<br>telefónicas y por correo.<br>«Cuadernos metodológicos<br>n.º 9». Madrid: Centro de<br>Investigaciones Sociológicas;<br>1993. | Charlton R. Research: is an «ideal»<br>questionnaire possible? Int   | Guyatt GH, Naylor CD, Juniper E,<br>Heyland DK, Jaeschke R, Cook DJ,<br>for the Evidence-Based<br>Medicine Working Group.<br>Users' guides to the medical<br>literature. XII. How to |

- use articles about health related quality of life. *JAMA*. 1997;277:1232-7.
- Hankins M. Questionnaire discrimination: (re)-introducing coefficient delta. *BMC Med Res Methodol*. 2007;7:19.
- McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. 2nd ed. New York: Oxford University Press; 1996.
- Navarro MD. Encuestas de salud. «Cuadernos metodológicos n.º 11». Madrid: Centro de Investigaciones Sociológicas; 1994.
- Rattray J, Jones MC. Essential elements of questionnaire design and development. *J Clin Nurs*. 2007;16(2):234-43.
- Streiner D, Norman G. *Health Measurement Scales – A practical guide to their development and use*. 3rd ed. revised. New York: Oxford University Press; 2003.

## Validación de cuestionarios

Antes de generalizar la aplicación de un cuestionario es necesario evaluar su fiabilidad y su validez, lo que supone tiempo y esfuerzo. Por ello, un investigador debería usar cuestionarios que ya hubieran mostrado su utilidad, fiabilidad y validez en otros estudios, lo que le permitiría además comparar los resultados.

Sin embargo, hay ocasiones en que es inevitable el uso de nuevos cuestionarios: cuando los existentes han mostrado resultados poco satisfactorios, cuando un cuestionario se ha mostrado eficaz pero en un medio distinto, o cuando no hay ninguno que sea adecuado para medir las variables de interés. En estas circunstancias, es preceptivo evaluar la utilidad del nuevo cuestionario a partir de dos criterios: su fiabilidad y su validez.

Los cuestionarios deben poseer una serie de características: ser sencillos, viables y aceptados (*feasibility*), fiables, válidos y bien adaptados culturalmente, útiles y sensibles a los cambios. Mientras que la fiabilidad y la validez son exigencias necesarias en todos los instrumentos, la importancia de otras características psicométricas dependerá del contexto; así, por ejemplo, la sensibilidad al cambio (*responsiveness*) será muy importante si el instrumento se aplica como medida de la respuesta en los ensayos clínicos, pero no lo será tanto en un estudio sobre opiniones o actitudes acerca de una enfermedad.

La validación de un cuestionario es un proceso complejo que implica múltiples fuentes de información y la recogida de diferentes evidencias empíricas. Dado que no existe un criterio concreto ni único a partir del cual pueda considerarse que un cuestionario es válido, en general es necesario realizar más de un estudio con esta finalidad. Las características consideradas en la validación de un cuestionario se detallan en el cuadro 21.1.

Cuadro 21.1 **Características a considerar en la validación de un cuestionario**

- Viabilidad
- Fiabilidad
  - Repetibilidad
  - Fiabilidad interobservador
  - Consistencia interna
- Sensibilidad
  - Sensibilidad al cambio
- Validez
  - Validez lógica
  - Validez de contenido
  - Validez de criterio
  - Validez de constructo o de concepto

### VIABILIDAD

Los mejores instrumentos son inservibles si su aplicación resulta compleja y costosa. Características como el tiempo empleado en la cumplimentación del cuestionario, la sencillez y la amenidad del formato, y el interés, la brevedad y la claridad de las preguntas, así como la facilidad de la puntuación, el registro y la codificación, y de la interpretación de los resultados, son algunos aspectos relacionados con la viabilidad.

### FIABILIDAD

Un instrumento es fiable si produce resultados consistentes cuando se aplica en diferentes ocasiones (estabilidad o reproducibilidad). Esquemáticamente,

se evalúa administrando el cuestionario a una misma muestra de sujetos, ya sea en dos ocasiones distintas (repetibilidad) y/o por dos observadores diferentes (fiabilidad interobservador). Se trata, por tanto, de analizar la concordancia entre los resultados obtenidos en las diversas aplicaciones del cuestionario. Si la escala de medida es cualitativa, se evalúa mediante el índice kappa, y si es cuantitativa, principalmente mediante el coeficiente de correlación intraclass (anexo 4). Otro concepto relacionado con la fiabilidad es el de la consistencia interna, que mide el grado en que se obtienen respuestas homogéneas a diferentes preguntas sobre un mismo concepto o dimensión. Siempre que sea posible deben evaluarse todos los componentes de la fiabilidad.

### Repetibilidad

La repetibilidad, o fiabilidad test-retest, se refiere a si, cuando se administra un cuestionario a la misma población en dos ocasiones diferentes en el tiempo, se obtienen resultados idénticos o similares.

Las limitaciones de la evaluación de la repetibilidad de un cuestionario son las mismas que las de cualquier otro instrumento de medida. Si el tiempo transcurrido entre ambas aplicaciones del cuestionario es muy largo, el fenómeno que se mide puede haber presentado variaciones, mientras que si es demasiado corto, puede existir un recuerdo de las respuestas dadas en la primera ocasión. En ambos casos se obtendrá una medida distorsionada de la repetibilidad. Además, algunos participantes pueden no aceptar que se administre el cuestionario en dos ocasiones, especialmente si es extenso.

**Ejemplo 21.1.** En la evaluación de la fiabilidad del cuestionario de salud Nottingham Impact Profile, uno de los aspectos que se tuvo en cuenta fue la repetibilidad (Hunt et al, 1981). El primer problema que se planteaba era la elección de la población. En estudios previos se había observado que las respuestas de los pacientes con enfermedades crónicas eran más fiables que las de los que padecían enfermedades agudas, y que las respuestas negativas eran más consistentes que las afirmativas. Dado que el Nottingham Impact Profile resulta en una alta proporción de respuestas negativas cuando se administra a una persona sana, una investigación sobre su fiabilidad se debería llevar a cabo en una población en la que se esperara un elevado número de respuestas afirmativas, con el fin de evitar su sobrevaloración. Por otro lado, para evitar una infraestimación de la repetibilidad, la muestra de la población debería padecer

una enfermedad estable y que previsiblemente no cambiara en un corto espacio de tiempo. Por estas razones, seleccionaron a pacientes con artrosis. Aunque el estado físico de una persona con artrosis puede fluctuar diariamente, es muy poco probable que se produzcan cambios significativos en un período de semanas.

Los autores consiguieron reclutar a 73 personas que cumplían los criterios de inclusión, y a cada una se le envió un cuestionario, una carta explicando el propósito del estudio y un sobre libre de franqueo para la respuesta. A todos aquellos que respondieron se les envió un segundo cuestionario 4 semanas más tarde. Este período se eligió para minimizar la posible sobrevaloración de la repetibilidad debida al efecto del recuerdo de las respuestas efectuadas en la primera ocasión. Los autores obtuvieron una tasa de respuesta del 88% en el primer cuestionario y del 90% en el segundo, francamente alta para este tipo de estudios, y la repetibilidad del cuestionario, a su juicio, fue buena. Sin embargo, se ha de recordar que, en el sentido más estricto, estos resultados sólo son aplicables a pacientes con artrosis. Muy posiblemente, los resultados sean extrapolables a otras poblaciones sanas o con otras enfermedades, aunque esto debe ser evaluado en cada ocasión.

### Fiabilidad interobservador

La evaluación de la fiabilidad interobservador consiste en estimar el grado de concordancia entre dos o más evaluadores (observadores). La demostración de una alta fiabilidad interobservador implica que la fiabilidad intraobservador también es alta. No obstante, si la fiabilidad interobservador es baja, no se puede asegurar si se debe a la existencia de diferencias entre los observadores o se debe a un solo observador.

**Ejemplo 21.2.** El Older Americans Resources and Services Multidimensional Functional Assessment Questionnaire (OARS-MFAQ) es un cuestionario sobre la capacidad funcional y las necesidades de atención de las personas de edad avanzada. Dado que las puntuaciones del cuestionario se basan en una revisión de las respuestas por un observador, la fiabilidad interobservador es especialmente importante. En un estudio (Fillenbaum y Smyer, 1981) se evaluó la concordancia entre 11 observadores que evaluaron a 30 pacientes, y se obtuvieron coeficientes de correlación intraclass que oscilaron entre 0,66 para el estado físico y 0,87 para los autocuidados.

## Consistencia interna

La consistencia interna se refiere a si los ítems que miden un mismo atributo presentan homogeneidad entre ellos.

Los cuestionarios se desarrollan para medir separadamente diferentes componentes o dimensiones de un problema. Un cuestionario de salud suele estar dividido en preguntas que tratan de medir la salud física y mental, o un cuestionario de satisfacción en apartados que identifican, por ejemplo, los componentes de competencia profesional, las cualidades personales del profesional sanitario y la accesibilidad de los servicios. En todas estas situaciones es de esperar que exista una buena homogeneidad entre las distintas preguntas que miden un mismo componente. Si en un cuestionario de satisfacción los usuarios contestan que su médico se preocupa bastante de ellos como persona, es de esperar que, en otra pregunta sobre si su médico está dispuesto a escucharlos, contesten afirmativamente. De otro modo, se pensará que los distintos ítems que componen la satisfacción con el médico que los atiende son poco consistentes entre sí y que el cuestionario es poco fiable.

Cuando un cuestionario está compuesto por diferentes subescalas, cada una de las cuales pretende medir una dimensión diferente del fenómeno, debe evaluarse la consistencia interna de cada una de ellas.

A diferencia de los otros aspectos de la fiabilidad, la evaluación de la consistencia interna sólo requiere la administración del cuestionario en una única ocasión.

La técnica estadística para su análisis es el alfa de Cronbach, que expresa la consistencia interna entre tres o más variables. Sus valores están comprendidos entre 0 y 1, y su interpretación es similar a la de un coeficiente de correlación. Pueden calcularse diferentes valores del alfa de Cronbach excluyendo determinados ítems del cuestionario, de forma que puede evaluarse si la supresión de algunas preguntas mejora la fiabilidad. De todas formas, antes de decidir eliminar un ítem, debe evaluarse si ello puede afectar a la validez del cuestionario, ya que puede ser preferible mantener la pregunta aun a costa de una consistencia interna ligeramente menor.

Como norma general, se sugiere que el valor del alfa de Cronbach ha de ser igual o superior a 0,70 para considerar que un instrumento tiene una buena consistencia interna.

**Ejemplo 21.3.** El cuestionario de apoyo social funcional Duke-UNC-11 consta de 11 ítems medidos en una escala de Likert con puntuaciones de 1 a 5,

y evalúa el apoyo social confidencial (posibilidad de contar con personas para comunicarse con ellas) y el afectivo (demostraciones de amor, cariño y empatía). En el marco de un estudio de su validez y fiabilidad, se evaluó también su consistencia interna (Bellón et al, 1996a). El alfa de Cronbach de la escala fue de 0,90, el de la subescala de apoyo confidencial de 0,88 y el del apoyo afectivo de 0,79. Los autores no observaron que la extracción de ningún ítem mejorara la consistencia interna de la escala ni de las subescalas.

## Fuentes de error

La fiabilidad de una medida puede afectarse por algunos de los factores que se citan a continuación:

- *Cambios a través del tiempo en la característica estudiada.* Al repetir un cuestionario se debe tener en cuenta qué medidas pueden variar con el tiempo. De hecho, muchas actitudes, creencias o estilos de vida pueden hacerlo, como la frecuencia de cepillado de los dientes o la práctica de ejercicio físico. Si ha existido un cambio, una repetibilidad baja no implica necesariamente una escasa fiabilidad del cuestionario.
- *Cambios debidos a las condiciones de administración del cuestionario.* Algunos factores personales del entrevistado, como el estado emocional, el cansancio, el estado de salud o las condiciones del entorno (ruido, calor, frío, etc.), pueden influir en el modo de contestar a las preguntas y alterar la fiabilidad.
- *Variaciones debidas al propio cuestionario.* En ocasiones, el formato del cuestionario, la formulación de las preguntas o las instrucciones para cumplimentarlo son poco comprensibles y pueden ser interpretadas de forma distinta por el encuestado al repetir la prueba. Por consiguiente, hay que insistir en que las instrucciones sean cortas y precisas y que las preguntas estén definidas de forma operativa con un lenguaje claro y sin ambigüedades.
- *Cambios atribuidos a los encuestadores.* Los encuestadores deben atenerse estrictamente a la estructura y la secuencia del cuestionario y a cómo han estado formuladas las preguntas. La transcripción de las respuestas a las preguntas abiertas debe ser literal. Al poner en marcha un estudio es de suma importancia entrenar a los entrevistadores para evitar estos errores y conseguir al mismo tiempo que exista uniformidad entre ellos.
- *Errores en el manejo de los datos.* Estos errores se pueden producir al codificar, grabar y/o transformar las variables para su análisis.

## VALIDEZ

La validez se refiere a la capacidad de un cuestionario para medir aquello para lo que ha sido diseñado; tiene diferentes aspectos o componentes, que deben ser evaluados en la medida de lo posible.

### Validez lógica

La validez lógica o aparente (*face validity*) es el grado en que parece que un cuestionario, una parte de él o un ítem mide lo que quiere medir.

La decisión sobre si las preguntas deben tener o no validez lógica ha de tomarse antes de iniciar su redacción. Si las preguntas carecen de validez lógica es muy probable que los encuestados rechacen contestar. De todos modos, en alguna ocasión puede ser de interés formular preguntas carentes de validez lógica. Por ejemplo, cuando se trata de temas muy sensibles, conflictivos o que no están bien vistos socialmente, si se realizan preguntas directas (con mucha validez lógica), es muy probable que el encuestado no responda o falsee la respuesta, por lo que puede ser preferible realizar preguntas que aborden el tema de una forma más indirecta, con una menor validez aparente.

### Validez de contenido

La validez de contenido se basa en el análisis del concepto que se pretende medir y, en especial, en la definición de las áreas o dimensiones que abarca y sus límites con otros conceptos relacionados. Se puede considerar que un cuestionario es válido por su contenido si contempla todos los aspectos relacionados con el concepto en estudio, lo que suele evaluarse a través de la opinión de expertos.

**Ejemplo 21.4.** Supongamos que se desarrolla un cuestionario para evaluar los conocimientos que tienen los estudiantes de medicina sobre una determinada enfermedad o sobre varios grupos de enfermedades. Para determinar la validez de contenido podría construirse una tabla en la que las columnas representarían las diferentes áreas de conocimiento (anatomía, fisiología, etiología, diagnóstico, etc.) y en cada columna se señalarían las preguntas del cuestionario que se incluyeran en dicha área. La simple inspección visual de la tabla permitiría apreciar si todas las áreas de conocimiento están representadas en la proporción deseada. El número de preguntas de cada área dependería de la importancia relativa de su contenido, habitualmente determinada por las opiniones de expertos e investigaciones previas sobre el tema.

Una forma empírica de evaluar la validez de contenido es aplicar un análisis factorial, técnica estadística que explora las respuestas a las preguntas del cuestionario, intentando agruparlas en función de factores subyacentes. Por ejemplo, si un cuestionario de estado de salud contiene dos subescalas, una que mide la salud física y otra la mental, es de esperar que el análisis detecte dos factores, cada uno de ellos formado por las preguntas que se relacionan con cada una de las subescalas. Para aplicar el análisis factorial, las escalas de medida deben ser cuantitativas o de puntuación por intervalos, y las respuestas han de seguir una distribución aproximadamente normal.

**Ejemplo 21.5.** En el estudio presentado en el ejemplo 21.3, en el que se evaluaba el cuestionario de apoyo social funcional Duke-UNC-11, que consta de 11 ítems medidos en una escala de Likert con puntuaciones de 1 a 5, se realizó un análisis factorial para detectar las dimensiones subyacentes en el cuestionario. El análisis reveló la existencia de dos factores: el primero con 7 ítems y el segundo con 4 ítems. Ambos factores explicaban, conjuntamente, el 60,9% de la variabilidad. Estos factores corresponden a los dos componentes teóricos del cuestionario original: el apoyo social confidencial (posibilidad de contar con personas para comunicarse con ellas) y el afectivo (demostraciones de amor, cariño y empatía).

La diferencia entre la validez aparente y la de contenido reside en que la evaluación de esta última es un proceso más exhaustivo, y quizás más formal, y en el que deberían participar tanto investigadores y clínicos como miembros de la población diana.

### Validez de criterio

En ocasiones se puede disponer de algún método alternativo de medida del fenómeno estudiado cuya validez haya sido demostrada, que se toma como referencia para determinar la validez de la encuesta. Siempre que se disponga de un método de referencia adecuado, deberá evaluarse la validez de criterio del cuestionario.

Cuando se habla de validar un cuestionario, los investigadores se suelen referir a la validez de criterio, que es, sin lugar a dudas, la más importante. En algunos casos se pueden usar como criterio de referencia medidas bioquímicas o radiológicas. Se puede validar, por ejemplo, el consumo de tabaco declarado, comparándolo con los valores derivados de la nicotina en sangre o de monóxido de carbono en el aire espirado. En otros casos, el investigador tendrá que fiarse de medidas menos objetivas, como

la historia clínica o los resultados obtenidos mediante otro cuestionario.

La validez de criterio puede evaluarse de dos formas: la validez concurrente y la validez predictiva. Para valorar la *validez concurrente* se relaciona la nueva medida con la de referencia, ambas administradas simultáneamente, de forma similar al estudio de la utilidad de una prueba diagnóstica. Cuando el criterio de referencia no esté disponible hasta un tiempo después (p. ej., el desarrollo de una enfermedad), se valora hasta qué punto la nueva medida es capaz de predecirlo correctamente, y se habla de *validez predictiva*. Cuando se evalúa la validez predictiva, los resultados del cuestionario no pueden influir sobre el criterio de referencia. En otras palabras, la evaluación de este criterio de referencia debe hacerse independientemente del resultado del cuestionario.

**Ejemplo 21.6.** Dado que la relación entre la capacidad pulmonar y la calidad de vida en enfermos pulmonares crónicos es débil, Guyatt et al (1987) desarrollaron un cuestionario sobre calidad de vida para utilizarlo en ensayos clínicos. Los existentes hasta el momento se centraban en la disnea y dejaban a un lado otros aspectos que pueden influir en la vida cotidiana de los pacientes. Por ello, desarrollaron un cuestionario que comprende cuatro grandes dimensiones: disnea, fatiga, estado emocional y la sensación de control del propio paciente sobre la enfermedad. La repetibilidad, que se evaluó en 100 pacientes con limitación crónica al flujo aéreo estable, fue excelente para las cuatro dimensiones. Los autores también evaluaron la validez de criterio, comparando las puntuaciones del cuestionario con los resultados espirométricos y con otros cuestionarios destinados a medir la disnea y la calidad de vida en general.

Cuando la escala de medida es cualitativa, los índices que se utilizan para evaluar la validez de criterio son la sensibilidad y la especificidad (anexo 3). Cuando se trata de una escala cuantitativa, se utiliza habitualmente el coeficiente de correlación intraclase (anexo 4).

A menudo surge el interrogante de por qué, si ya existe un buen criterio de referencia, interesa una nueva medida. El desarrollo de esta nueva medida está justificado si el criterio de referencia es muy caro, requiere mucho tiempo de administración, presenta muchos efectos secundarios, o bien no se desarrolla hasta un tiempo después. En las tres primeras situaciones el interés se centra en evaluar la validez concurrente, mientras que en la última interesa determinar la validez predictiva del cuestionario.

## Validez de constructo o de concepto

A veces resulta imposible evaluar la validez de criterio, ya que éste no existe o no está al alcance del investigador. En estos casos, el procedimiento más empleado es evaluar la validez de constructo, que engloba distintas estrategias. La *validez discriminante* se refiere a la capacidad para distinguir entre subgrupos de pacientes o individuos con distintos niveles del atributo de interés. Por ejemplo, es de esperar que la calidad de vida relacionada con la salud será peor en los pacientes asmáticos con gran afectación funcional que en aquellos con formas más leves.

El método más sencillo para evaluar la validez discriminante es el de los *grupos extremos*, que consiste en administrar el cuestionario a dos grupos de sujetos: uno con la característica o conducta de interés, y otro que carece de ella. Este enfoque presenta dos problemas. El primero es la propia definición de los grupos extremos, ya que no siempre existe un criterio adecuado para conocer quién tiene y quién no tiene la característica de interés. En este caso se puede dividir la muestra en función de la puntuación obtenida con el propio instrumento, seleccionando, por ejemplo, el 30% de los sujetos con mejores puntuaciones y el 30% de los que tienen las peores puntuaciones. El segundo problema es similar al que se presenta al evaluar una prueba diagnóstica: puede ser relativamente sencillo discriminar entre dos grupos muy extremos, pero ésta no es la utilidad que se pretende dar al instrumento en la práctica habitual. Por tanto, comprobar que un cuestionario es útil para diferenciar entre dos grupos extremos no es suficiente para demostrar su validez.

**Ejemplo 21.7.** En un estudio que tenía por objetivo evaluar la validez y fiabilidad de un cuestionario de función familiar (Bellón et al, 1996b), se utilizó el método de los grupos extremos para evaluar la validez de constructo. Se partió de la hipótesis de que los casados tienen una mejor función familiar que los divorciados, por lo que deberían obtener puntuaciones superiores en el cuestionario.

Otra estrategia para evaluar la validez de constructo es comprobar que el cuestionario se correlaciona con otras variables que se cree que están relacionadas con él (*validez convergente*), mientras que no lo hace con otras con las que se sospecha que no tiene relación alguna (*validez divergente*).

**Ejemplo 21.8.** En un estudio se evaluó la validez y la fiabilidad del cuestionario de apoyo social funcional Duke-UNC-11 en una muestra de 656 pacientes

(Bellón et al., 1996a). Para evaluar la validez de constructo, los autores eligieron determinadas características que, según la literatura científica, están relacionadas con el apoyo social, y que son la edad, el estado civil, el estado de salud, la utilización de servicios, la salud mental, la función familiar y el número de convivientes. Para ello, compararon las distribuciones de estas variables entre los sujetos con un apoyo social normal o bajo, según el cuestionario, y determinaron los coeficientes de correlación entre el cuestionario evaluado y el resto de las escalas cuantitativas.

**Ejemplo 21.9.** Para validar el test de Fresno, un cuestionario que mide los conocimientos y habilidades de los profesionales sanitarios en medicina basada en la evidencia (Argimon et al., 2010), los autores seleccionaron tres grupos en función de su experiencia previa. El primero estaba formado por profesionales expertos que habitualmente participaban en el diseño y la realización de ensayos clínicos, el segundo por tutores de medicina de familia y el tercero por médicos residentes. La hipótesis subyacente era que la puntuación del test sería mayor en los grupos más expertos comparado con el de los residentes. El grupo con profesionales más expertos obtuvo una puntuación de 149,8 sobre un máximo de 212 puntos, mientras que los residentes sólo obtuvieron 60,4 puntos. Los tutores de medicina de familia obtuvieron una puntuación intermedia (110,4 puntos), lo que confirmó la hipótesis de los investigadores.

## SENSIBILIDAD AL CAMBIO

El concepto de sensibilidad (*sensitivity*) se refiere a la capacidad de un instrumento para detectar casos verdaderos.

La sensibilidad al cambio (*responsiveness*) es la capacidad que tiene un instrumento para detectar cambios clínicos importantes en el estado de salud de los participantes (o en el atributo que se mide) a lo largo del tiempo, y puede considerarse un tipo especial de validez de constructo o concepto. La sensibilidad al cambio es un aspecto crucial en los ensayos clínicos, la valoración de programas y los análisis de coste-utilidad; en otras palabras, cuando el instrumento se utiliza como variable de respuesta.

Su conocimiento también es importante para la estimación del cálculo del tamaño muestral necesario para demostrar los cambios debidos al tratamiento en un ensayo clínico. Cuanto mayor sea la

sensibilidad al cambio del instrumento, menor será el tamaño de la muestra necesario.

La sensibilidad al cambio se estima en estudios longitudinales, y existen dos abordajes para calcularla. El primero se basa en la distribución estadística y evalúa los cambios que se producen en la puntuación de la escala y su variabilidad asociada, generalmente la desviación estándar. Existen varios índices que la miden, todos ellos muy parecidos. Los índices más usados son el tamaño del efecto (*effect size*), la media del cambio estandarizada (*standardised response mean*) y el índice de sensibilidad al cambio. El tamaño del efecto es la diferencia entre los valores basales y los del final del estudio dividida por la desviación estándar de los valores basales. Su resultado refleja el cambio estandarizado en el número de desviaciones estándar de los valores basales. Como norma general, el tamaño del efecto puede considerarse pequeño si es menor de 0,2, pequeño-moderado si se encuentra entre 0,2 y 0,5, moderado-grande si va de 0,51 a 0,79, y grande si supera el valor de 0,79. La media del cambio estandarizada se calcula como la media del cambio en las puntuaciones, antes y después, dividida por la desviación estándar de estos cambios. Los valores positivos reflejan incrementos estandarizados en el número de desviaciones estándar de las diferencias de puntuaciones. El índice de sensibilidad al cambio se define como la diferencia media entre los valores basales y los obtenidos al final del seguimiento, dividida por la desviación estándar de los individuos más estables (p. ej., aquellos que han recibido un placebo). Este índice es el que suele mostrar valores más elevados, ya que la desviación estándar es más pequeña.

La ventaja de estos índices estadísticos es que son fáciles de calcular en comparación con el método de ancla. Sin embargo, tienen dos limitaciones importantes: la estimación de la variabilidad difiere de un estudio a otro y su interpretación no es muy intuitiva para el clínico. Además, los cambios «estadísticamente significativos» dependen del número de participantes en el estudio y, por tanto, no equivalen necesariamente a un cambio clínicamente importante, o viceversa.

**Ejemplo 21.10.** En un estudio se incluyeron 152 médicos residentes y se analizaron los cambios en sus conocimientos sobre medicina basada en la evidencia, medidos con el test de Fresno, antes y después de recibir un seminario (Argimon et al., 2011). El tamaño del efecto fue de 1,77 (intervalo de confianza del 95% [IC 95%]: 1,57-1,95) y la respuesta media estandarizada fue de 1,65 (IC 95%: 1,47-1,82). Las diferencias entre ambos índices,

aunque no son muy grandes, se explican porque persiguen objetivos distintos. Mientras que los dos índices usan la diferencia entre las puntuaciones obtenidas antes y después de la intervención, el estimador de la variabilidad es distinto. La respuesta media estandarizada usa la desviación estándar de las diferencias y su objetivo es mostrar que existen cambios estadísticamente significativos, mientras que el tamaño del efecto usa la desviación estándar antes de la intervención y su objetivo es cuantificar la magnitud del efecto.

En el segundo abordaje, denominado el método de ancla, las puntuaciones de la escala que se valida se comparan con los valores de otra variable o escala validada, con el fin de observar si las puntuaciones de los dos instrumentos se correlacionan. La sensibilidad al cambio estimada por este método siempre es más sencilla de interpretar. Lo que es muy importante es que la variable que se usa como ancla sea por sí misma una medida válida del cambio clínico que puedan experimentar los participantes del estudio. Su principal limitación es que no tiene en cuenta ninguna medida de la precisión de la variable, y de este modo no se puede descartar que un cambio relevante en las puntuaciones sea debido a una gran variabilidad de la medida.

Para que una escala pueda emplearse como variable de respuestas en un ensayo clínico, es necesario que los ítems que la formen midan todos los conceptos relevantes de la enfermedad o del problema que se esté estudiando, y que la mayoría de ellos sean capaces de detectar cambios. La sensibilidad al cambio vendrá determinada por la interacción entre los propios ítems que formen la escala, la intervención que se evalúe y la población en la que se aplique.

Si al contestar un ítem de un cuestionario los pacientes no reflejan tener un problema de salud, entonces este ítem no será válido para detectar una mejoría. Del mismo modo, si un ítem sólo se altera en los casos más graves, en los estudios donde se incluyan pacientes menos graves este ítem tampoco será útil para detectar cambios. Por otro lado, si un ítem se altera en todos los pacientes porque es una característica de la enfermedad, tampoco servirá para detectar cambios, a menos que la intervención que se evalúe consiga la curación del paciente. En definitiva, los ítems que muestren un efecto «techo» (*ceiling effect*) o un efecto «suelo» (*floor effect*) no serán sensibles para detectar cambios.

Los efectos «techo» o «suelo» dependen, en parte, de la población de estudio. Un ítem puede mostrar un efecto «cielo» cuando se estudia a pacientes muy graves, pero no cuando estos pacientes tienen un

grado menor de afectación del estado de salud. Los efectos «cielo» o «techo» se pueden inferir de los resultados de las pruebas piloto o de los estudios que se hayan efectuado en distintas poblaciones.

Estos efectos también dependen de la validez de contenido del cuestionario. Si un ítem es irrelevante para un grupo de población, su capacidad para detectar un cambio a lo largo del estudio será pequeña.

**Ejemplo 21.11.** Los resultados de algunas investigaciones sobre el asma sugieren que los ítems relacionados con la práctica de un deporte son más sensibles al cambio en las personas jóvenes, mientras que en las mayores lo son menos. Una explicación es que, para las personas mayores, la posibilidad de practicar un deporte es menos importante. La relevancia de un ítem depende mucho de la población en la que se administra el cuestionario.

La relevancia de un ítem se manifiesta al comparar escalas específicas de una enfermedad o un problema de salud para medir la calidad de vida o el estado de salud percibido por los pacientes con escalas genéricas. Como norma general, las escalas específicas suelen contener más ítems con capacidad para detectar cambios que las genéricas y, por consiguiente, son las más usadas en los ensayos clínicos.

El que un ítem sea o no sensible a los cambios también depende del formato de respuesta. En general, las respuestas tipo escala de Likert son las más sensibles para detectar cambios. Cuantos más puntos haya en las categorías, más sensible será el instrumento a los cambios.

## Fuentes de error

Los aspectos del diseño del cuestionario que pueden influir en su validez son los siguientes:

- *Orden de las preguntas.* Es conveniente situar las más conflictivas al final del cuestionario, ya que si se ponen al principio, existe la posibilidad de que la persona encuestada rechace seguir respondiendo o no lo haga con la sinceridad deseada.
- *Redacción de las preguntas,* ya que puede inducir una respuesta sesgada.

**Ejemplo 21.12.** Supongamos una pregunta en la que se plantea la posibilidad de que una enfermedad afecte a 600 individuos y se informa de que existen dos programas alternativos: el programa A, que salvará 200 vidas, y el B, con el que hay un tercio de probabilidades de salvarse y dos tercios de morir. La gran mayoría de los encuestados preferirá

el programa A. Nótese que los resultados de los programas A y B (en términos de vidas salvadas) son los mismos. La diferencia está en cómo se ha explicado o descrito la situación. En el programa A, el número de supervivientes queda explícito, mientras que el número que muere (400) queda implícito. Desde un punto de vista aritmético, el número de sujetos que debería preferir uno u otro programa tendría que ser el mismo o similar. Esta aparente paradoja se puede explicar porque los individuos tienen aversión al riesgo cuando se presenta una situación de posible ganancia y, por el contrario, aceptan el riesgo cuando la situación se presenta desde el punto de vista de pérdida.

- *Errores en la categorización de las respuestas.* En los cuestionarios con opciones de respuestas predeterminadas hay que tener en cuenta todas las posibilidades de respuesta, de modo que el encuestado encuentre siempre una opción adecuada. En otras ocasiones, son los factores personales del entrevistado los que introducen los sesgos en las respuestas. Hay individuos que, al ser encuestados, tienden a contestar de forma distinta a como lo harían normalmente.
- *Sesgo de conveniencia social.* Algunas personas tienden a dar la respuesta socialmente más aceptable o la que piensan que contesta la mayoría de los individuos. Este sesgo depende de muchos factores individuales, como la edad, el sexo, la clase social del entrevistado y el contexto en el que se realiza la pregunta. Ejemplos de ello se pueden encontrar en cuestionarios o preguntas sobre el consumo de alcohol o el aborto. Un mecanismo para intentar evitarlo es dar en primer lugar las opciones menos convenientes, de modo que sea más fácil seleccionarlas. En muchas ocasiones, el sesgo de conveniencia social se produce sin que el entrevistado tenga intención de engañar, mientras que en otras el individuo no dice lo que piensa de forma deliberada.
- *Tendencia sistemática a dar siempre la misma respuesta.* Otra fuente de error es la tendencia sistemática de algunos individuos a contestar «sí», «verdadero» o «de acuerdo» a las cuestiones planteadas. En el caso más extremo, estas respuestas se dan independientemente del contenido de la pregunta. Como ejemplo se podría dar el caso de un individuo que contestara afirmativamente a la pregunta de si toma medicación todos los días a la hora indicada, y también lo hiciera cuando se le preguntara si a menudo se olvida de tomar la medicación. En el otro extremo del espectro se encuentran los que siempre contestan negativamente o no están de acuerdo.

- *Características del encuestador.* La respuesta de una persona a la pregunta de si cumple con la medicación prescrita puede ser muy distinta si la realizan los profesionales sanitarios que atienden al paciente o si lo hace un entrevistador no implicado directamente en su seguimiento médico.
- *Elección de la categoría o puntuación intermedia.* En las escalas de puntuación o las de Likert existe la posibilidad, además, de que aparezcan otros sesgos. Uno de ellos se refiere a que algunos individuos siempre escogen la categoría o puntuación intermedia. El efecto de este sesgo es la reducción, en la práctica, de las posibles respuestas. Así, si en una escala de Likert con cinco opciones, las extremas recogen la posibilidad de «siempre» o «nunca», un *sesgo de aversión hacia los extremos* implica que la escala queda reducida a tres categorías, ya que el individuo no contesta ni «nunca» ni «siempre», con la consiguiente pérdida de fiabilidad. Existen dos posibles soluciones a este problema. La primera es evitar términos absolutos y, en su lugar, utilizar «casi siempre» y «casi nunca». La segunda consiste en aumentar las categorías posibles en cada respuesta; así, si en realidad se desea una pregunta con cinco categorías de respuesta, finalmente tendrá siete, de modo que las categorías extremas sean las ocupadas por los términos absolutos.

## ADAPTACIÓN TRANSCULTURAL

La adaptación de un cuestionario a otra cultura tiene como objetivo conseguir un instrumento equivalente al desarrollado en el país de origen. No puede limitarse a una simple traducción, sino que debe seguir una metodología que asegure la equivalencia conceptual y semántica con el original y la comprensión por parte de los pacientes de la versión adaptada. El método más utilizado es el de la traducción-retraducción por personas bilingües (fig. 21.1), seguido de un análisis de la nueva versión para detectar discrepancias, y de la comprobación de su comprensión y aceptabilidad en un grupo de pacientes.

Para realizar una adaptación transcultural, en primer lugar debe evaluarse si el fenómeno que tiene que medir el cuestionario existe en la cultura a la que se desea adaptar. Por ejemplo, si el cuestionario trata sobre cómo una enfermedad o un problema de salud afecta al grado o la manera de realizar una determinada actividad, hay que preguntarse si estos conceptos se operativizan de igual modo en una cultura que en otra. Es esencial conocer qué significado tiene el concepto y cómo se expresa en



**Figura 21.1** Esquema del proceso de adaptación transcultural de un cuestionario.

la cultura original para poder identificar similitudes en la nuestra.

A continuación, hay que traducir el cuestionario. Deben realizarse al menos dos traducciones conceptuales más que literales. La persona que traduce el cuestionario no sólo debe ser un perfecto conocedor de los dos idiomas, sino también de los contenidos y los propósitos del cuestionario, ya que la traducción literal de un término puede tener un significado distinto según el idioma. Por ejemplo, en inglés se usa la palabra *blue* para designar tristeza. Si la frase: «*I feel blue*» se tradujera literalmente al español no tendría ningún sentido. A partir de estas traducciones, el equipo investigador consensúa una primera versión del cuestionario. Es conveniente que un panel de expertos evalúe la equivalencia conceptual de esta versión con la original.

Después se retraduce el cuestionario dos veces al idioma original por al menos dos individuos bilingües, evaluando su equivalencia conceptual con el cuestionario original. Es conveniente realizar una revisión del cuestionario por un panel de población general o de pacientes de diferentes características sociodemográficas con la finalidad de llegar a un acuerdo sobre los ítems que se van a incluir. A partir de toda esta información se consensúa una versión preliminar. Finalmente, debe realizarse un estudio

piloto de esta versión en una muestra de sujetos de características similares a las de la población a la que se administrará el cuestionario, para poder evaluar su comprensión y viabilidad.

Una vez obtenida la versión final, deben comprobarse sus propiedades métricas mediante el reescalamiento de los ítems y dimensiones en nuestro medio, ya que su importancia, o valor relativo, puede de ser distinta en ambas culturas. Y, finalmente, debe comprobarse su validez y fiabilidad. Además, si pretende utilizarse para medir la respuesta en estudios longitudinales, hay que evaluar su sensibilidad al cambio.

**Ejemplo 21.13.** Un estudio tenía por objetivo adaptar el cuestionario genérico de estado de salud SF-36 Health Survey para su uso en España (Alonso et al, 1995). Este proyecto se estaba realizando de forma coordinada en diferentes países. El cuestionario original fue traducido al castellano por dos personas bilingües con experiencia clínica. Ambas traducciones fueron discutidas simultáneamente entre los traductores y un miembro del equipo investigador hasta alcanzar un consenso. Otras dos personas bilingües evaluaron la equivalencia conceptual, la claridad y la naturalidad de cada una de las frases y de las opciones de respuesta de esta primera versión. A continuación fue traducido al inglés por dos personas bilingües. Sus dos retraducciones fueron comparadas con la versión original por un equipo de expertos, quienes señalaron algunos ítems o palabras que no parecían tener una completa equivalencia conceptual con el original. También se realizó una reunión con los autores de todas las versiones del cuestionario existentes en diferentes países, durante la cual se trató de armonizar el contenido del cuestionario en los casos en que existían diferentes expresiones alternativas del mismo concepto. Por ejemplo, se sustituyó la distancia de una milla por la de un kilómetro. Se realizaron diferentes estudios piloto con diversos grupos de pacientes crónicos para valorar la comprensibilidad del cuestionario y la factibilidad de su administración.

## SELECCIÓN Y USO DE UN INSTRUMENTO DE MEDIDA

Una vez validados, los instrumentos de medida se usan generalmente con dos fines distintos: para evaluar la efectividad de una intervención sanitaria o para describir y/o discriminar entre grupos de pacientes.

Las escalas que se usan en los ensayos clínicos para valorar la efectividad han de contener una elevada proporción de ítems sensibles al cambio, no suelen ser muy largas y, habitualmente, no superan los 30 o 40 ítems.

Por el contrario, cuando un instrumento se emplea en un estudio transversal con el fin de describir el estado de salud y/o discriminar entre los pacientes, no es necesario que el número de ítems sea reducido. Un instrumento tendrá más capacidad de discriminación cuanto mayor sea el número de ítems que contenga.

En un ensayo clínico no conviene usar instrumentos que contengan ítems frente a los que más del 70% de los pacientes opte por la respuesta más extrema (efecto «techo» o «cielo»). Sin embargo, en un estudio transversal interesa usar instrumentos donde los pacientes opten por todas las posibles respuestas en función de su gravedad. Por tanto, los pacientes más graves o con menos calidad de vida optarán por una respuesta extrema, mientras que los que tengan menor afectación optarán por otra respuesta.

La elección entre una escala genérica o específica para una enfermedad dependerá del objetivo del estudio.

**Ejemplo 21.14.** Supóngase que en un estudio se emplea un instrumento genérico para valorar la calidad de vida de un grupo de pacientes diagnosticados de enfermedad pulmonar obstructiva crónica (EPOC), y que este instrumento contiene un número importante de ítems relacionados con el dolor. Como los pacientes con EPOC suelen ser mayores y presentan otras enfermedades, algunas de ellas cuyo síntoma principal puede ser el dolor, existirá una gran variación en las puntuaciones de estos enfermos. Supóngase que se desea, a continuación, correlacionar la puntuación obtenida en el cuestionario de calidad de vida con una medida de la función pulmonar. Debido a la gran variabilidad observada en la puntuación del cuestionario de calidad de vida, es posible que se reduzca la posible asociación entre calidad de vida y la medida de la función pulmonar. En este caso sería preferible el uso de un cuestionario específico, en el que no existieran interferencias de preguntas que no tuvieran una relación directa con la enfermedad. Por el contrario, si se deseara una medida general del estado

de salud de una población, o compararlo con el de otro grupo, teniendo en cuenta la comorbilidad, sería preferible una medida genérica.

La utilidad de un instrumento depende básicamente de su fiabilidad y validez, pero también de su interpretabilidad clínica. Ésta viene dada por el grado en que se pueden realizar juicios de valor sobre un resultado cuantitativo, que permitan, por ejemplo, la toma de decisiones clínicas. La estrategia más utilizada para aumentar la interpretabilidad de los instrumentos de medida de la salud percibida ha sido su administración a una muestra representativa de la población general para obtener valores o normas poblacionales de referencia, basados habitualmente en el cálculo de los percentiles.

Cuando el instrumento se usa como variable de respuesta, además de la sensibilidad al cambio es importante disponer del valor que puede considerarse como el cambio mínimo que puede estimarse como relevante desde el punto de vista clínico. El método más recomendado es usar varios marcadores clínicos o anclas, junto con el tamaño del efecto o la respuesta media estandarizada como información de soporte, para finalizar obteniendo el valor o rango de valores considerados relevantes desde el punto de vista clínico.

En ausencia de una medida objetiva o ancla, la diferencia mínima de interés clínico puede estimarse como una proporción la desviación estándar del tamaño del efecto.

**Ejemplo 21.15.** Norman et al (2003) realizaron una revisión sistemática de 38 estudios (que incluían 62 tamaños del efecto) y observaron, con sólo unas pocas excepciones, que las mínimas diferencias clínicamente relevantes se correspondían con la mitad de la desviación estándar del tamaño del efecto.

Es preferible usar la desviación estándar de la puntuación basal del tamaño del efecto antes que la desviación estándar de los cambios antes y después, como en la respuesta media estandarizada, ya que el propósito es decidir la magnitud del cambio, no su significación estadística.

De todos modos, la diferencia mínima no debe considerarse un valor fijo, sino que depende del contexto y puede variar de un estándar a otro y entre poblaciones.

**BIBLIOGRAFÍA DE LOS EJEMPLOS**

- 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;104:771-6.
- Argimon-Pallàs JM, Flores-Mateo G, Jiménez-Villa J, Pujol-Ribera E. Effectiveness of a short-course in improving knowledge and skills on evidence-based practice. *BMC Fam Pract*. 2011;12:64.
- Argimon-Pallàs JM, Flores-Mateo G, Jiménez-Villa J, Pujol-Ribera E. Psychometric properties of a test in evidence based practice: the Spanish version of the Fresno test. *BMC Med Educ*. 2010;10:45.
- Bellón JA, Delgado A, Luna del Castillo JD, Lardelli P. Validez y fiabilidad del cuestionario de apoyo social funcional Duke-UNC-11. *Aten Primaria*. 1996;18:153-63.
- Bellón JA, Delgado A, Luna del Castillo JD, Lardelli P. Validez y fiabilidad del cuestionario de función familiar Apgar-familiar. *Aten Primaria*. 1996;18:289-96.
- Fillenbaum GG, Smyer MA. The development, validity and reliability of the OARS Multidimensional Functional Assessment Questionnaire. *J Gerontol*. 1981;36:428-34.
- Guyatt G, Berman L, Townsend M, Pugsley S, Chambers L. A measure of quality of life for clinical trials in chronic lung disease. *Thorax*. 1987;42:773-8.
- Hunt S, Mc Kenna, Williams J. Reliability of a population survey tool for measuring perceived health problems: a study of patients with osteoarthritis. *J Epidemiol Community Health*. 1981;35:297-300.
- Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Med Care*. 2003;41:582-92.

**BIBLIOGRAFÍA**

- Altman DG, Bland JM. Cronbach's alpha. *BMJ*. 1997;314:572.
- Badia X, Salamero M, Alonso J. La medida de la salud: guía de escalas de medición en español. 3.<sup>a</sup> ed. Barcelona: Edimac; 2002.
- Bland JM, Altman DG. Validating scales and indexes. *BMJ*. 2002;324:606-7.
- Cohen RD. Validation of health-related quality of life instruments. *Hepatology*. 1999;29(6 Suppl):7-8.
- Epstein RS. Responsiveness in quality-of-life assessment: nomenclature, determinants, and clinical applications. *Med Care*. 2000;38(9 Suppl):1191-4.
- Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc*. 2002;77:371-83.
- Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *Int J Chron Obstruct Pulmon Dis*. 2005;2:63-7.
- Herdman M. La medida de la calidad de vida relacionada con la salud. *Med Clin (Barc)*. 2000;114(Suppl 3):22-5.
- Hyland ME. A brief guide to the selection of quality of life instrument. *Health Qual Life Outcomes*. 2003;1(1):24.
- Knauper B, Turner PA. Measuring health: improving the validity of health assessments. *Qual Life Res*. 2003;12(Suppl 1):81-9.
- Liang MH, Lew RA, Stucki G, Fortin PR, Daltroy L. Measuring clinically important changes with patient-oriented questionnaires. *Med Care*. 2002;40(4 Suppl):II45-51.
- Moran LA, Guyatt GH, Norman GR. Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. *J Clin Epidemiol*. 2001;54:571-9.
- Patrick DL, Chiang YP. Measurement of health outcomes in treatment effectiveness evaluations: conceptual and methodological challenges. *Med Care*. 2000;38(9 Suppl):II14-25.
- Revicki D, Hays RD, Celli D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102-9.
- Revicki DA, Celli D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes*. 2006;4:70.
- Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc Sci Med*. 1999;48(11):1531-48.
- Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. 3rd ed. New York: Oxford Medical Publications; 2003.
- Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res*. 2003;12:349-62.

## Estrategia de análisis

La planificación de la estrategia de análisis forma parte del diseño del estudio y no debe dejarse para después de haberse recogido los datos. Para planificar adecuadamente la estrategia es necesario tener presente el objetivo específico del estudio y el marco conceptual que permite establecer las interrelaciones entre las diferentes variables, lo que ayuda a definir cuál es el papel de cada una de ellas en el análisis.

El esquema que se presenta en este capítulo no es rígido, y debe ser adaptado a las características concretas de cada estudio. Es importante, sin embargo, mantener su línea general, que podría resumirse en ir de lo sencillo a lo complicado.

En este capítulo se exponen los pasos que seguir en la estrategia de análisis, y se reservan para el próximo los problemas relacionados con la decisión de qué individuos deben ser incluidos en él.

### REVISIÓN DE LOS DATOS

Al finalizar la recogida de datos, es frecuente que el investigador esté impaciente por responder a la pregunta planteada, y desee realizar rápidamente los pasos finales del análisis. El problema es que, si no se ha asegurado la calidad de los datos, los resultados pueden ser erróneos. Por ello, una fase previa al análisis será la revisión de la matriz de datos.

Hay que prever un análisis de las variables una a una, en busca de valores no habituales o ilógicos, o de errores de transcripción o codificación. Para esta finalidad, son útiles las técnicas de presentación y síntesis de datos que componen la estadística descriptiva. Una tabla de frecuencias permitirá detectar posibles anomalías, como errores de codificación o transcripción, que obligarán a revisar la hoja de recogida de datos para realizar las correcciones. Al

mismo tiempo, podrá comprobarse si la distribución de frecuencias es similar a la esperada, de forma que podrán detectarse errores que de otra forma pasarían inadvertidos.

La distribución de frecuencias puede indicar la necesidad de realizar agrupaciones o desagregaciones de determinados valores. Si el estado civil se ha codificado en múltiples categorías (soltero, casado, viudo, separado, divorciado), puede ocurrir que en algunas de ellas exista un reducido número de casos que dificulte el uso y la interpretación de las pruebas estadísticas. En este caso, puede resultar adecuado realizar una agrupación de valores. O bien, puede darse la situación contraria. Si una categoría es poco específica y contiene un alto porcentaje de casos, puede resultar conveniente desagregarla en categorías más pequeñas que puedan ser interpretadas adecuadamente. Por ejemplo, si se recoge información sobre enfermedades concomitantes y se codifica como hipertensión arterial, diabetes y otras, puede ocurrir que la categoría «otras» incluya un importante número de casos. En esta situación sería conveniente desagregar esta categoría.

En el caso de las variables cuantitativas, hay que prever revisar su distribución y no conformarse simplemente con determinar su media y desviación estándar. Si esta última es muy elevada en relación con la media, indica una gran dispersión de los valores o una falta de normalidad en la distribución, lo que deberá tenerse en cuenta para seleccionar la prueba estadística más adecuada. También resulta muy útil observar con detenimiento la distribución de los valores, ya sea gráficamente o utilizando un número suficiente de percentiles. De esta forma, pueden detectarse valores extremadamente altos o bajos, que pueden influir sobre la media, alterando los resultados de las pruebas estadísticas. Estos

valores extremos deben ser verificados y corregidos si es necesario. Por ejemplo, valores de potasemia de 10 mEq/l son incorrectos con toda seguridad, ya que son incompatibles con la vida, o valores de colesterolémia muy bajos (p. ej., inferiores a 125 mg/dl) pueden requerir una confirmación, dada su escasa probabilidad. Si los criterios de inclusión fijan un rango de edad entre los 15 y los 45 años, valores fuera de este rango pueden corresponder a la inclusión errónea de algún sujeto o a errores en la transcripción del dato.

Al mismo tiempo, si se desea reclasificar una variable cuantitativa en una cualitativa mediante la creación de intervalos, para facilitar la comprensión de los resultados, el conocimiento de la distribución de frecuencias puede ser de inestimable ayuda para definir sus límites.

También conviene prever cruces de variables dos a dos para detectar valores improbables o ilógicos. Por ejemplo, los varones no pueden estar tomando anticonceptivos orales o tener antecedentes de hysterectomía. Es improbable que un sujeto de 16 años tenga 5 hijos, por lo que debe verificarse si el dato es correcto.

Mediante estos sencillos procedimientos, y otros similares, pueden corregirse gran parte de los errores de transcripción y codificación, aumentando la calidad de los datos de la matriz en la que se basará el análisis.

## ESTRATEGIA DE ANÁLISIS DE UN ESTUDIO DESCRIPTIVO

La estrategia que se presenta es útil para los estudios cuyo objetivo es estimar un parámetro (p. ej., la prevalencia o la incidencia de un problema de salud) en la población ([cuadro 22.1](#)). Las estrategias útiles para los estudios con otros objetivos se comentan al presentar sus diseños específicos.

**Cuadro 22.1 Pasos de la estrategia de análisis de un estudio descriptivo**

0. Revisar los datos previamente al análisis
1. Describir los sujetos estudiados
2. Evaluar la representatividad de la muestra
3. Estimar el valor del parámetro de interés
4. Calcular el intervalo de confianza de la estimación
5. Describir las no respuestas y/o pérdidas de seguimiento y evaluar su impacto potencial
6. Análisis de subgrupos

## Describir los sujetos estudiados

La descripción de la muestra de los sujetos estudiados permite:

- Evaluar la adecuación de la muestra respecto al problema planteado.
- Conocer en qué tipo de sujetos se ha obtenido el resultado y poder interpretar correctamente las conclusiones del estudio.
- Evaluar la posibilidad de extrapolar los resultados del estudio a otras poblaciones. Los lectores podrán determinar las diferencias entre las características de su población y la estudiada, y tomar en consideración cómo pueden influir sobre los resultados.
- Replicar el estudio por otros investigadores, utilizando criterios similares, de forma que puedan detectarse diferencias entre poblaciones, ámbitos geográficos, etc.

No se trata de describir exhaustivamente la muestra, sino de sintetizar sus principales características relacionadas con las finalidades citadas ([cuadro 22.2](#)).

**Cuadro 22.2 Características de la población estudiada que deben ser descritas en el análisis\***

1. Fuente de los sujetos
2. Mecanismo de selección e inclusión de los grupos
3. Características sociodemográficas
4. Características que permiten evaluar la aplicabilidad del protocolo (criterios de inclusión y exclusión)
5. Características de la enfermedad estudiada y su tratamiento, o de la exposición y su grado:
  - Criterios diagnósticos
  - Tiempo de evolución de la enfermedad
  - Gravedad de la enfermedad
  - Presencia de complicaciones
  - Tipo de tratamiento
  - Otras características
6. Situación basal: punto de partida (en estudios longitudinales)
7. Factores que pueden influir en el resultado
8. Otras características de interés (si no están incluidas en los apartados anteriores)

\* Las categorías no son excluyentes y algunas características pueden situarse en más de una simultáneamente

## Evaluar la representatividad de la muestra

Debe preverse qué variables se analizarán para evaluar la representatividad de la muestra estudiada en relación con la población de estudio. Si, por ejemplo, ésta la forman individuos de 20 a 70 años de edad, residentes en una comunidad, y la muestra se extrae del padrón municipal, la representatividad se estudiará comparando la distribución de la edad de los sujetos de la muestra con la del padrón municipal de toda la población candidata.

## Estimar el valor del parámetro de interés

La medida de la frecuencia del problema de salud debe ser la adecuada al objetivo y al diseño del estudio (anexo 1).

Un diseño transversal permite estimar la prevalencia, es decir, la proporción de sujetos de la muestra que presentan la característica o el problema de salud estudiado, que puede ser de dos tipos: la *prevalencia de punto, puntual o instantánea*, que corresponde a la proporción o la frecuencia relativa de individuos de la muestra que presentan una característica en un momento determinado, y la *prevalencia de período*, o proporción de sujetos que ha presentado el fenómeno o la característica en algún momento durante el período considerado.

Cuando el problema de salud es de carácter prolongado, y tiene interés determinar si está presente o ausente en cada elemento de una muestra en un instante dado (p. ej., hipertensión arterial, diabetes mellitus, esquizofrenia o desempleo), la medida de elección es la prevalencia puntual. En cambio, cuando se trata de procesos agudos o de fenómenos que cursan con episodios más o menos breves, como la migraña, la lumbalgia aguda o la gripe, puede ser preferible la prevalencia de período.

La prevalencia de período plantea el problema de determinar el valor del denominador, ya que el número de elementos de la población puede haber cambiado a lo largo del tiempo considerado y hay que decidir en qué momento se hará la estimación correspondiente (habitualmente la del punto medio del período).

Dependiendo de la rareza de la característica considerada, la proporción suele presentarse multiplicada por 100 o por  $10^n$ . Por ejemplo, la prevalencia de hipertensión arterial en personas mayores de 65 años podría expresarse como porcentaje, pero la de esclerosis múltiple en la población general probablemente habría de darse en tanto por millón o por 100.000 habitantes.

Los estudios longitudinales permiten estimar la frecuencia de aparición de un problema de salud, denominada *incidencia*. Existen dos tipos: la *incidencia acumulada*, que es la proporción de individuos sanos que desarrolla la enfermedad a lo largo de un período determinado, y la *densidad de incidencia*, que se define como el número de nuevos casos que ocurren por unidad de persona-tiempo en riesgo. La elección del más adecuado viene condicionada por el objetivo del estudio (anexo 1). Si el interés está en predecir el cambio en el estado de salud de un individuo en función de alguna característica, entonces se deberá estimar el riesgo de que este cambio ocurra (incidencia acumulada). Si el tiempo de seguimiento es muy desigual entre los sujetos de la muestra, deberá optarse por la densidad de incidencia.

## Calcular el intervalo de confianza de la estimación

Sea cual sea la medida utilizada, la estimación puntual debe acompañarse de su correspondiente *intervalo de confianza* (IC). El valor observado en la muestra no es más que uno de los posibles valores que hubieran podido obtenerse si se hubieran estudiado las múltiples muestras que pueden extraerse de una población, cada una de las cuales conduciría a un resultado diferente debido a la variabilidad del muestreo. Por ello, se necesita una medida de esta variabilidad que permita conocer entre qué límites se espera que se encuentre el verdadero valor de la población con una confianza determinada, lo que se hace mediante el IC.

El IC proporciona mucha más información que la simple estimación puntual, ya que permite conocer la precisión con que el parámetro poblacional ha sido estimado, es decir, cuál es el posible error de la estimación puntual en relación con el verdadero pero desconocido valor, debido a la variabilidad aleatoria. Habitualmente, el IC se calcula para un nivel de confianza del 95%.

## Describir las no respuestas y/o pérdidas de seguimiento y evaluar su impacto potencial

La descripción de las no respuestas y pérdidas deberá incluir no sólo su número, sino también sus motivos, ya que pueden suponer un sesgo. Para evaluar su impacto potencial pueden ser útiles diferentes estrategias de análisis:

- Comparar los sujetos que responden y los que no lo hacen, por las variables disponibles. Por ejemplo,

si se desea seleccionar, a partir del padrón municipal, una muestra de individuos de ambos sexos, con edades comprendidas entre los 20 y los 70 años, debería conocerse la edad y el sexo de los que responden y de los que no lo hacen. Si se observan diferencias entre ellos, las no respuestas podrían introducir un sesgo de selección.

- *Comparar los que responden a la primera llamada de los investigadores con los que lo hacen en llamadas sucesivas.* En este caso se podrá comparar no sólo por las variables utilizadas para seleccionar la muestra, sino también por la frecuencia de la exposición o de la enfermedad que se está estudiando. La asunción que se realiza en esta situación es que aquellos que responden más tardíamente tienen características intermedias entre los que responden rápidamente y los que no responden.

## Análisis de subgrupos

Cuando los investigadores piensan que la frecuencia del problema de salud puede variar en función de diferentes características, puede ser práctico planificar la estimación del parámetro en los diferentes subgrupos, justificando el interés de dicho análisis.

Si se desea realizar un análisis de subgrupos, deberá tenerse en cuenta en el cálculo del tamaño de la muestra, para evitar la pérdida de precisión que se produciría como consecuencia del menor número de sujetos de los subgrupos, así como en el método de selección de los sujetos, para garantizar una adecuada representatividad.

## ESTRATEGIA DE ANÁLISIS DE UN ESTUDIO ANALÍTICO

El objetivo de los estudios analíticos es estimar la magnitud del efecto o la asociación entre un factor de estudio y una variable de respuesta. La finalidad de la estrategia de análisis es obtener la estimación más válida de este parámetro, eliminando la influencia de todas las variables que puedan interferir, y lo más precisa posible. Las fases de la estrategia se resumen en el cuadro 22.3.

### Describir los sujetos estudiados

Su utilidad y las características por las que se realiza esta descripción son similares a las comentadas para los estudios descriptivos (ver cuadro 22.2).

**Cuadro 22.3 Pasos de la estrategia de análisis de un estudio analítico**

0. Revisar los datos
1. Describir los sujetos estudiados
2. Evaluar la comparabilidad inicial de los grupos
3. Estimar la magnitud del efecto o asociación:
  - Elegir la medida del efecto
  - Evaluar la existencia de modificación del efecto
  - Ajustar por potenciales factores de confusión
  - Decidir las pruebas estadísticas
4. Calcular el intervalo de confianza de la estimación
5. Análisis de subgrupos
6. Responder las preguntas secundarias

## Evaluar la comparabilidad inicial de los grupos

El análisis se basa en que los grupos de estudio son comparables y no existen sesgos que alteren los resultados. Por ello, antes de determinar si el efecto o la asociación es diferente entre los grupos, debe planificarse un análisis de su comparabilidad basal. Si se encuentran diferencias relevantes en alguna variable importante, deberá controlarse su potencial efecto de confusión en fases posteriores del análisis.

El análisis de esta comparabilidad no se basa en la significación estadística, ya que algunas diferencias podrían ser estadísticamente significativas sólo por azar, mientras que diferencias relevantes podrían no serlo si el número de sujetos es reducido. Además, los grupos podrían diferir en alguna característica importante, que no hubiera sido medida en el estudio.

**Ejemplo 22.1.** En un ensayo clínico aleatorio a doble ciego cuyo objetivo era comparar la eficacia del trimetoprim-sulfametoxazol en pauta de 10 días y en monodosis, en mujeres con sintomatología de disuria aguda (Fihn et al, 1988), se incluyó a 255 mujeres que cumplían los criterios de selección; 126 de ellas recibieron la monodosis, y 129, la pauta de 10 días. En la tabla 22.1 se presentan los resultados de la evaluación de la comparabilidad inicial de ambos grupos respecto a las variables pronósticas de la respuesta. Puede comprobarse que son comparables para todas las variables analizadas, excepto en el uso de diafragma, más frecuente en el grupo asignado a la monodosis. Si se considera que el uso de diafragma puede influir sobre el efecto del tratamiento, la

**Tabla 22.1** Características de las 255 mujeres incluidas en un ensayo clínico aleatorio comparativo de la eficacia de trimetoprim-sulfametoxazol en monodosis respecto a la pauta de 10 días (ejemplo 22.1)

Características de las pacientes	Grupos	
	Monodosis (n = 126)	10 días (n = 129)
Edad (media ± DE)	24,5 ± 3,5	24,5 ± 4,2
Casadas (%)	4	9
Infecciones urinarias previas:		
Antes de los 12 años (%)	12	11
Últimas 6 semanas (%)	9	10
Últimos 6 meses (%)	31	29
Últimos 12 meses (%)	44	40
Número de episodios (media ± DE)	3,9 ± 5,2	3,6 ± 4,3
Número de parejas sexuales masculinas en los últimos 6 meses (media ± DE)	1,7 ± 1,1	1,5 ± 1,0
Frecuencia de relaciones sexuales al mes (media ± DE)	12,6 ± 10,2	12,3 ± 7,8
Usuarias de diafragma (%)*	64	48
Días de duración de los síntomas (media ± DE)	4,2 ± 4,9	4,0 ± 4,8

DE: desviación estándar.

\*p=0,013.

diferencia en su uso entre los grupos deberá tenerse en cuenta en fases posteriores del análisis.

## Estimar la magnitud del efecto o de la asociación

La planificación de esta fase del análisis supone elegir la medida óptima del resultado, identificar las variables que pueden influir sobre el efecto de interés y utilizar las pruebas estadísticas adecuadas.

## Elección de la medida del efecto

La medida más adecuada depende del tipo de estudio y su finalidad, y de la escala de medida de las variables. Existen dos tipos de medidas: relativas y absolutas. Las *medidas relativas* más utilizadas son el *rriesgo relativo* (RR) y la *odds ratio* (OR) (anexo 2).

El RR corresponde al cociente entre las incidencias observadas en el grupo expuesto y en el no expuesto, y por tanto, sólo puede calcularse directamente en aquellos estudios que permitan determinar incidencias (estudios de cohortes y ensayos clínicos). El RR indica el número de veces que es más frecuente la aparición de la respuesta en el grupo de estudio respecto al de referencia. Por este motivo, al preparar la estrategia de análisis es importante decidir cuál

será la categoría que se utilizará como referencia, y definirla con claridad.

En los estudios de casos y controles no puede calcularse directamente el RR, por lo que se utiliza la OR. Si el estudio está bien diseñado y se cumplen determinadas asunciones, la OR es una buena estimación del RR.

Cada vez es más frecuente el uso de la OR como medida relativa del efecto en investigación clínica y epidemiológica, debido a que puede calcularse independientemente del diseño del estudio. Además, las técnicas estadísticas multivariantes de uso habitual, como la regresión logística o el modelo de Cox, facilitan su cálculo.

Otra forma de expresar los resultados de un estudio es a través de *medidas absolutas*. La más utilizada es la *diferencia de riesgos*, o diferencia entre las incidencias observadas en ambos grupos, que proporciona una idea del impacto que tendría sobre la incidencia la eliminación de un factor de riesgo o la aplicación de una intervención. En los estudios de casos y controles no puede determinarse directamente, ya que estos estudios no permiten estimar la incidencia.

**Ejemplo 22.2.** En el estudio sobre la disuria aguda del ejemplo 22.1, la respuesta sintomática fue uno

de los criterios de evaluación del estudio. Se midió en tres tiempos diferentes. Se observó que las diferencias favorecían al grupo tratado con pauta de 10 días. Así, por ejemplo, en el grupo tratado con monodosis, el 56,1% de las mujeres no presentaba síntomas a las 6 semanas, frente a un 63,6% de las mujeres del grupo tratado con pauta de 10 días. La diferencia de los porcentajes fue, pues, del 7,5% (IC 95%: -5,5 a 20,6). Además de medir el efecto en la escala absoluta como diferencia en las respuestas observadas, los autores también estimaron el efecto en una escala relativa ( $OR = 1,4$ ; IC 95%: 0,8 a 2,4).

Las medidas que se utilizan para cuantificar el efecto de una intervención sanitaria se presentan en el anexo 5.

### Variables que controlar

Al planificar la estrategia de análisis, es importante identificar todas las variables implicadas a partir del marco conceptual elaborado al diseñar el estudio, y decidir por cuáles deberá ajustarse la estimación, por cuáles no, y qué factores podrían modificar el efecto estudiado. En esta fase resulta muy útil una buena revisión bibliográfica que permita identificar los posibles factores de confusión y modificadores del efecto tenidos en cuenta por otros investigadores.

Según su función en la relación causa-efecto que se esté estudiando, pueden distinguirse los siguientes tipos de variables (fig. 22.1):

- **Variable de respuesta.** En términos estadísticos, suele denominarse *variable dependiente*.
- **Factor de estudio.** Suele denominarse *variable independiente principal*, ya que la finalidad del análisis es estimar su efecto sobre la variable dependiente, ajustando o controlando la influencia del resto de las variables.
- **Variables modificadoras del efecto.** Cuando el efecto o la asociación entre el factor de estudio y la variable de respuesta varían según la existencia

de una tercera variable o de diferentes niveles de ella, se dice que ésta es modificadora de dicho efecto. Es importante identificar estas variables, ya que, más que controlar su influencia, interesa describir cómo modifican el efecto que se está estudiando.

**Ejemplo 22.3.** En el estudio sobre el tratamiento de la disuria aguda (ejemplo 22.1) se observó que los resultados en las mujeres cuyo cultivo inicial presentaba menos de 100.000 colonias/ml eran diferentes en relación con los resultados de las que presentaban 100.000 colonias/ml o más. Por lo tanto, el número de colonias en el cultivo inicial es una variable modificadora del efecto del tratamiento sobre la respuesta al mismo, por lo que deberán presentarse los resultados separadamente en función del número de colonias.

- **Potenciales factores de confusión.** Son variables que están relacionadas tanto con el factor de estudio como con la variable de respuesta, de forma que, si no se tienen en cuenta en el análisis, se obtiene una estimación distorsionada del efecto. Por lo tanto, deberá ajustarse la estimación del efecto por estas variables.

**Ejemplo 22.4.** Supóngase un estudio que desea estimar el efecto del consumo de alcohol sobre el riesgo de presentar una enfermedad coronaria. Un potencial factor de confusión es el consumo de tabaco, ya que se sabe que las personas que consumen más alcohol fuman con más frecuencia, y que los fumadores tienen un mayor riesgo de enfermedad coronaria. Si no se ajusta por el consumo de tabaco, se obtendrá una estimación sesgada del efecto del consumo de alcohol sobre el riesgo de enfermedad coronaria.

- **Variables de la cadena causal.** Se relacionan con el factor de estudio y con la variable de respuesta, pero se encuentran situadas en la cadena causal,

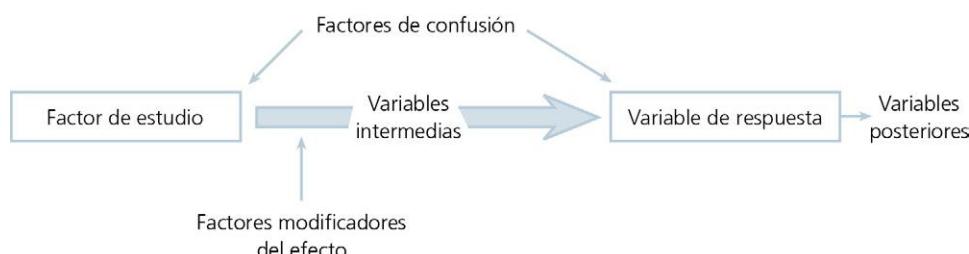


Figura 22.1 Papel de las variables en la estrategia de análisis.

ya sea antes (variables intermedias) o después de la variable de respuesta (variables posteriores). En principio no debe ajustarse por este tipo de variables, ya que, al formar parte de la cadena causal, la estimación ajustada del efecto del factor de estudio no tendría en cuenta el componente que pasa a través de las variables intermedias o el relacionado con las variables posteriores.

**Ejemplo 22.5.** Supóngase un estudio que trata de estimar el efecto del consumo de tabaco durante el embarazo sobre la mortalidad perinatal. Una variable intermedia podría ser el bajo peso al nacer, ya que puede ser consecuencia del consumo de tabaco y, a la vez, es un factor de riesgo de mortalidad perinatal. A diferencia del ejemplo anterior, la variable intermedia no sólo está asociada al factor de estudio, sino que es consecuencia del mismo en la cadena causal que conduce hacia la respuesta. Si se elimina el efecto debido al bajo peso al nacer (es decir, si se ajusta por esta variable), se obtendrá una estimación sesgada del efecto del consumo de tabaco, que no tendrá en cuenta la parte del efecto que se produce a través del bajo peso al nacer.

**Ejemplo 22.6.** Supóngase un estudio que trata de estimar el efecto del consumo de tabaco durante el embarazo sobre la probabilidad de bajo peso al nacer. En esta situación, la mortalidad perinatal sería una variable posterior y no debería ajustarse por ella, ya que se eliminaría del análisis el efecto del consumo de tabaco sobre la mortalidad perinatal, de forma que sólo se estimaría el efecto del consumo de tabaco sobre el riesgo de bajo peso al nacer no relacionado con la mortalidad.

### Elección de la prueba estadística

En la planificación de la estrategia de análisis es necesario prever las pruebas estadísticas que se utilizarán, que dependerán fundamentalmente del tipo de variables implicadas, el número de grupos que se compararán, el tamaño de la muestra y las condiciones de aplicación específicas de cada una de las pruebas (anexo 9).

Si es necesario obtener una estimación del efecto, ajustando por múltiples variables, debe recurrirse a modelos estadísticos multivariantes. La elección del modelo apropiado depende de la escala de medida de la variable de respuesta (variable dependiente) y de otras condiciones de aplicación específicas. Los modelos más utilizados en investigación clínica y epidemiológica son la regresión lineal múltiple (cuando la variable de respuesta es cuantitativa), la regresión logística (cuando es dicotómica) y el

modelo de Cox (cuando la variable de respuesta es el tiempo de aparición de un suceso).

### Confusión y modificación del efecto

La estrategia de análisis ha de planificarse de forma que permita obtener la mejor estimación posible de la magnitud del efecto del factor de estudio sobre la variable de respuesta, teniendo en cuenta las influencias del resto de variables que pueden interferir.

En primer lugar, deben identificarse las variables que pueden ser modificadoras del efecto, prever la estimación del efecto en los diferentes estratos definidos por los valores de dichas variables, y la aplicación de pruebas estadísticas para comparar las estimaciones obtenidas. Si se detecta la existencia de modificación del efecto, deberá tenerse en cuenta al presentar los resultados, ya que no existirá un único resultado, sino que variará en función de las categorías o los valores de otras variables. Si no se detecta ninguna modificación del efecto, deberá obtenerse una estimación del resultado ajustada por los potenciales factores de confusión.

**Ejemplo 22.7.** En el estudio sobre el tratamiento de la disuria aguda del ejemplo 22.1, se analizaron los resultados en función de diferentes variables. En la tabla 22.2 se presentan los resultados en función del número de colonias en el cultivo inicial. Se observa que las OR de los estratos son muy distintas entre sí, lo cual sugiere que el efecto de la monodosis respecto a la pauta de 10 días produce resultados diferentes en función del recuento inicial de colonias. Si se calculara una única OR ajustada, se obtendría el valor 1,60, que resumiría inadecuadamente los resultados, ya que parecería indicar que el porcentaje de fallos con la monodosis es superior en todos los sujetos, cuando el análisis estratificado indica que ello es así en las pacientes con más de 100.000 colonias/ml ( $OR = 1,97$ ), mientras que parece ser todo lo contrario si el recuento inicial de colonias es menor ( $OR = 0,63$ ). Por ello, cuando existe modificación de efecto, no es adecuado resumir los resultados en un solo índice ponderado (OR ajustada), sino que deben presentarse separadamente para cada estrato. De todas formas, para poder asegurar que existe modificación de efecto debería comprobarse que la diferencia de los valores de OR de cada estrato es estadísticamente significativa, aunque si el número de sujetos de los estratos es reducido, esta prueba puede no tener la potencia suficiente para detectarla.

**Ejemplo 22.8.** Siguiendo con el mismo estudio del ejemplo 22.1, se había detectado una diferencia estadísticamente significativa en el porcentaje de

**Tabla 22.2** Fallos del tratamiento con trimetoprim-sulfametoaxazol (monodosis frente a pauta de 10 días) en mujeres con disuria aguda, en función del número de colonias del cultivo inicial (ejemplo 22.7)

	Grupo		OR (IC 95%)
	Monodosis (fallos/casos)	10 días (fallos/casos)	
Total	29/90	20/97	1,8 (0,9 a 3,5)
Estratos (cultivo inicial)			
< 100.000 colonias/ml	2/19	6/38	0,6 (0,1 a 3,4)
≥ 100.000 colonias/ml	27/71	14/59	2,0 (0,9 a 4,3)

OR: odds ratio.

IC 95%: intervalo de confianza del 95%.

**Tabla 22.3** Fallos del tratamiento con trimetoprim-sulfametoaxazol (monodosis frente a pauta de 10 días) a los 42 días en mujeres con disuria aguda, en función del uso de diafragma (ejemplo 22.8)

	Grupo		OR (IC 95%)
	Monodosis (fallos/casos)	10 días (fallos/casos)	
Total	29/108	20/111	1,7 (0,9 a 3,2)
Estratos (uso de diafragma)			
Sí	20/66	13/54	1,4 (0,6 a 3,1)
No	9/42	7/57	1,9 (0,7 a 5,7)

OR: odds ratio.

IC 95%: intervalo de confianza del 95%.

mujeres que utilizaban diafragma entre los grupos de estudio. Dado que esta variable puede estar asociada con la respuesta terapéutica, por ejemplo, a través de la infección por patógenos diferentes, podría interferir en la medición de la respuesta. Por ello, puede recurrirse al análisis estratificado, analizando por separado las mujeres que lo utilizan y las que no, y evaluando las diferencias en la respuesta al tratamiento en cada grupo (tabla 22.3). Se aprecia que en el grupo tratado con monodosis existe un mayor porcentaje de mujeres que utilizan diafragma (el 61 frente al 48%), y que éstas tienen un mayor porcentaje de fallos terapéuticos que las que no lo utilizan (el 27,5 frente al 16%). Por lo tanto, puede cuestionarse si la mayor tasa de fracasos observada en el grupo de monodosis puede ser consecuencia del mayor porcentaje de mujeres que utilizan diafragma incluidas en dicho grupo. Los datos de la tabla muestran que, en ambos estratos, la tasa de fracasos fue más elevada en el grupo tratado con monodosis, si bien la diferencia respecto al otro grupo fue ligeramente menor en el estrato de las mujeres que utilizan diafragma (OR = 1,4 frente a 1,9).

Dado que la diferencia entre las OR de los estratos no es estadísticamente significativa, se considera que no existe modificación de efecto y puede calcularse una medida global del efecto eliminando la influencia de la posible variable de confusión, de modo que se obtiene una OR ajustada de 1,6 (IC 95%: 0,7 a 3,3).

Los dos ejemplos anteriores han evaluado la influencia de dos variables separadamente. Sin embargo, la estrategia de análisis debe diseñarse de forma que permita obtener una estimación del efecto que tenga en cuenta todas las variables simultáneamente, y no una a una. Para lograrlo, el análisis estratificado es poco eficiente, ya que podría ser necesario utilizar gran cantidad de estratos, lo que dificultaría la interpretación de los resultados, además de reducir el número de sujetos en cada uno de ellos, lo que haría que los procedimientos estadísticos fueran poco potentes. Por ello, es preferible utilizar el análisis multivariante que obvia estas limitaciones. En el capítulo 30 se comentan más detalladamente los conceptos de factor de confusión y modificación de efecto y las técnicas de análisis disponibles para su evaluación.

## Determinar la precisión de la estimación del efecto

Una vez calculada la estimación que se considere más válida, deberá acompañarse de su correspondiente IC como medida de la precisión.

## Análisis de subgrupos

El análisis de subgrupos se realiza cuando interesa evaluar los resultados en algún grupo especial de sujetos, habitualmente en función de su edad, sexo o alguna variable pronóstica.

El análisis por subgrupos se justifica si se ha planteado antes del inicio del estudio, en función de una hipótesis fundamentada. En algunas ocasiones, aunque no haya sido establecido previamente, el análisis de los resultados observados en algún subgrupo de sujetos puede ser razonable. Sin embargo, no hay que olvidar que, si se define un gran número de subgrupos en función de diferentes combinaciones de variables, un análisis indiscriminado puede conducir a la obtención de conclusiones erróneas, al aparecer el problema de las comparaciones múltiples, por lo que es preferible planificar a priori los análisis de subgrupos que se realizarán y justificarlos adecuadamente. El cuadro 22.4 recoge los criterios para considerar que los resultados de un análisis de subgrupos son válidos.

## Preguntas secundarias

A menudo, interesa responder a preguntas secundarias. Por ejemplo, si se va a realizar un estudio sobre la eficacia de un programa de educación sanitaria

### Cuadro 22.4 Criterios de validez de un análisis de subgrupos

- Hipótesis específica sobre los resultados del subgrupo planteada a priori en el diseño del estudio como una de las pocas hipótesis que se querían contrastar
- Análisis de subgrupos realizado dentro de una estrategia adecuada de comparaciones múltiples
- Magnitud del efecto clínicamente relevante
- Efecto estadísticamente muy significativo (cuanto más bajo es el valor de  $p$ , más creíble es la diferencia)
- Resultados consistentes con los de otros estudios
- Evidencia indirecta que avale los resultados encontrados (plausibilidad biológica)

sobre el control metabólico de los pacientes diabéticos, también interesaría responder a preguntas tales como: ¿qué grupo de población ha asistido a las sesiones?, ¿qué características socioculturales definen a los no cumplidores?, ¿qué individuos han incrementado su nivel de conocimientos sobre la enfermedad?, etc. Las respuestas pueden ayudar a enriquecer la discusión y la interpretación de los resultados. Sin embargo, están sometidas a los mismos inconvenientes que se han citado para el análisis por subgrupos, por lo que deben ser escasas en número, establecidas a priori y estar debidamente fundamentadas. En su interpretación, deben recordarse las limitaciones de las comparaciones múltiples.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

Fihn SD, Johnson C, Roberts PL, Running K, Stamm WE. Trimetoprim-sulfamethoxazole for acute dysuria in women: a single-dose or 10-day course. A double blind randomized trial. Ann Intern Med. 1988;108:350-7.

## BIBLIOGRAFÍA

- |  |  |  |
|--|--|--|
| Altman DG, Bland JM. Interaction revisited: the difference between two estimates. BMJ. 2003;326:219. | intervals illuminate absence of evidence. BMJ. 2004;328:1016-7.                                      | Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet. 2000;355:1064-9. |
| Altman DG, Bland JM. Units of analysis. BMJ. 1997;314:1874.  | Medidas del efecto de un tratamiento. Atención Primaria Basada en la Evidencia. 1998;5(Supl 7):45-6. | Barraclough H, Govindan R. Biostatistics primer: what  |
| Altman D, Bland JM. Confidence   |  |  |

- a clinician ought to know: subgroup analyses. *J Thorac Oncol.* 2010;5:741-6.
- Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ.* 1995;310:170.
- Cobo E, Buekens P. Necesidad y limitaciones del ajuste. *Med Clin (Barc).* 1990;95:702-8.
- DeMets DL. Statistical issues in interpreting clinical trials. *J Intern Med.* 2004;255:529-37.
- Fletcher J. Subgroup analyses: how to avoid being misled. *BMJ.* 2007;335:96-7.
- Goodman SN. Multiple comparisons, explained. *Am J Epidemiol.* 1998;147:807-12.
- Hemilä H, Kaprio J. Subgroup analysis of large trials can guide further research: a case study of vitamin E and pneumonia. *Clin Epidemiol.* 2011;3:51-9.
- Hernández AV, Boersma E, Murray GD, Habbema JD, Steyerberg EW. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? *Am Heart J.* 2006;151:257-64.
- Lang JM, Rothman KJ, Cann CI. That confounded P-value. *Epidemiology.* 1998;9:7-8.
- Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med.* 1992;116:78-84.
- Ramalle Górnara E, Valídez de los análisis de subgrupos. *FMC.* 1999;6(Supl 2):1-2.
- Savitz DA, Olshan AF. Multiple comparison and related issues in the interpretation of epidemiological data.
- Am J Epidemiol. 1995;142:904-8.
- Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ.* 2010;340:c117.
- Tripepi G, Jager KJ, Dekker FW, Zoccali C. Measures of effect in epidemiological research. *Nephron Clin Pract.* 2010;115:c91-3.
- Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. *BMJ.* 2001;323:1123-4.
- Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA.* 1991;266:93-8.

## Sujetos a incluir en el análisis

En el capítulo anterior se ha presentado el esquema de la estrategia del análisis de los datos, asumiendo que se disponía de la información completa de todos los sujetos. Sin embargo, es muy frecuente que no hayan podido medirse todas las variables en todos los sujetos, e incluso suelen existir casos en los que no ha podido determinarse la variable de respuesta. Además, pueden presentarse ciertas situaciones que obliguen a considerar si determinados sujetos u observaciones deben ser excluidos del análisis. Según la actitud que se adopte ante estas situaciones, las conclusiones del estudio pueden ser diferentes.

La exclusión de sujetos u observaciones del análisis puede plantear dos tipos de problemas. Por un lado, disminuye la potencia estadística, ya que el número de individuos u observaciones que se tiene en cuenta es inferior al inicialmente previsto, por lo que existirá una menor capacidad para detectar la diferencia o asociación de interés. Este problema se previene aumentando el tamaño de la muestra en función del porcentaje esperado de abandonos o pérdidas. Por otro lado, estas situaciones especiales pueden vulnerar los principios de representatividad y comparabilidad de los grupos si los motivos por los que determinados sujetos no son considerados en el análisis son diferentes, introduciendo un sesgo y comprometiendo la validez interna del estudio. El incremento del tamaño de la muestra no evita este segundo problema, por lo que es fundamental intentar que no se produzca este tipo de situaciones. Las situaciones especiales que obligan a considerar la posible exclusión de determinados casos u observaciones del análisis pueden ser diferentes según el diseño del estudio (*cuadro 23.1*). No existe una clasificación ni una terminología unánimemente aceptada, por lo que es recomendable que el protocolo del estudio describa claramente estas

**Cuadro 23.1 Situaciones que pueden plantear la posible exclusión de determinados sujetos del análisis de un ensayo clínico aleatorio**

- Violaciones de los criterios de inclusión y exclusión
- Diagnóstico incorrecto
- Asignación incorrecta
- Falta de cooperación o voluntad de continuar en el ensayo
- Pérdida de contacto durante el seguimiento
- Enfermedades intercurrentes
- Insuficiente cumplimiento con la intervención asignada
- Administración de una intervención diferente de la asignada
- Uso de medicación concomitante prohibida
- Fallecimiento por causas no relacionadas con la enfermedad en estudio
- Ausencia de datos
- No finalización del estudio por acontecimientos adversos
- No finalización del estudio por falta de eficacia

situaciones y defina qué debe hacerse cuando se presentan. De la misma forma, el informe final del estudio debe explicar las circunstancias por las que determinados sujetos no han finalizado el estudio o han sido excluidos del análisis, el momento en que se ha producido el problema y la información de que se dispone, para poder evaluar las posibles implicaciones que tienen sobre la interpretación y extrapolación de los resultados.

En este capítulo se describen el impacto y el abordaje de las no respuestas, las pérdidas de seguimiento, las retiradas y, más específicamente para los ensayos clínicos, las exclusiones antes de la asignación aleatoria.

## NO RESPUESTAS

Habitualmente, cuando se habla de no respuestas se hace referencia a los estudios transversales, en los que se ha seleccionado una muestra de sujetos de los que va a recogerse una determinada información. Las no respuestas corresponden a los sujetos que forman parte de esta muestra, pero que no colaboran en el estudio o de los que no se obtiene la información deseada, de forma que el análisis únicamente del conjunto de sujetos de los que se dispone de información puede vulnerar el principio de representatividad.

Conceptualmente, el problema de las no respuestas radica no tanto en su número o en el porcentaje que suponen respecto a la muestra, sino en los motivos por los que se han producido. Si están asociadas al problema en estudio o a determinadas características relacionadas con éste, el conjunto de sujetos analizados no es representativo de la población que se deseaba estudiar, y el resultado obtenido no será una buena estimación del verdadero valor de dicha población.

Si, por el contrario, las no respuestas se producen de forma aleatoria o se deben a motivos que no están relacionados con el problema en estudio, la consecuencia será que la estimación que se obtenga tendrá una menor precisión, o la comparación, una menor potencia estadística, ya que se habrá estudiado un número menor de individuos, pero no estará sesgada.

En realidad, el impacto potencial de las no respuestas sobre la medida del fenómeno de interés depende tanto de su número como de la frecuencia de presentación de dicho fenómeno.

**Ejemplo 23.1.** Supongamos un estudio transversal que desea estimar la prevalencia de una enfermedad a partir de una muestra de 200 sujetos, de los que 20 rehúsan participar. De los 180 restantes, 36 presentan la enfermedad, por lo que la estimación puntual de la prevalencia en ellos es del 20%. Si ninguno de los 20 sujetos que han rehusado participar tuviera la enfermedad, la verdadera prevalencia en la muestra de 200 sujetos sería del 18% ( $36/200 = 0,18$ ). Por el contrario, si los 20 sujetos tuvieran la enfermedad, sería del 28% ( $[36 + 20]/200 = 0,28$ ). Si las no

respuestas fueran aleatorias y se comportaran como el resto de los sujetos de la muestra, cabría esperar que 4 de ellos tuvieran la enfermedad, por lo que la prevalencia sería del 20% ( $[36 + 4]/200 = 0,20$ ), por lo que la estimación de la prevalencia no habría sido sesgada.

Supongamos que en este mismo estudio, de los 180 sujetos que participan, tan sólo 4 tuvieran la enfermedad, de forma que la estimación puntual de la prevalencia sería del 2,2%. Si ninguno de los 20 sujetos que no participan tuviera la enfermedad, la prevalencia sería del 2% ( $4/200 = 0,02$ ), mientras que, si los 20 la tuvieran, sería del 12% ( $[4 + 20]/200 = 0,12$ ). En este caso, en que la prevalencia de la enfermedad es inferior, el impacto potencial de las no respuestas puede ser mayor, ya que puede haber estimaciones hasta seis veces superiores.

Finalmente, si en lugar de 20 no respuestas se hubieran producido tan sólo 2, su impacto potencial sobre las estimaciones sería mucho menor. Y si hubieran sido 100 las no respuestas, su impacto podría ser todavía mayor que el descrito.

El diseño del estudio debe prever mecanismos que ayuden a conseguir el máximo número de respuestas y a recaptar las no respuestas cuando se producen. Así mismo, en el análisis de los datos es conveniente estudiarlas y evaluar si el valor de la variable de interés difiere en relación con el de las respuestas. Por ello, es frecuente que, si se han realizado varios intentos para captar a los sujetos de la muestra, se analicen por separado los resultados de cada intento para ver si existen diferencias en la estimación del valor de la variable de interés, y poder así evaluar un posible sesgo. Es importante describir las causas por las que se han producido las no respuestas y evaluar si existen diferencias, en relación con las respuestas, en las características que puedan estar relacionadas con la variable de interés.

También en los estudios analíticos puede aparecer el problema de las no respuestas. En los estudios de casos y controles suelen ser más frecuentes en el grupo de los controles, en especial cuando provienen de la población general. Estas situaciones pueden alterar la estimación de la frecuencia de la exposición en los grupos, de forma que, si su impacto es diferente en los casos y en los controles, además del principio de representatividad, puede vulnerarse también el de comparabilidad.

En los estudios de cohortes también puede afectarse el principio de comparabilidad, ya que las no respuestas se producen cuando los grupos ya están formados (dado que los sujetos no son asignados a los grupos, sino que forman parte de ellos por

determinadas características sobre las que los investigadores no tienen control).

**Ejemplo 23.2.** Supongamos que se realiza un estudio de cohortes para evaluar el efecto del consumo de tabaco sobre la incidencia de cardiopatía isquémica. Supongamos que todos los no fumadores aceptan participar, ya que son sujetos preocupados por su salud, mientras que, entre los fumadores, existe un porcentaje de sujetos que rehúsa participar tanto más elevado cuanto mayor es su consumo de tabaco y peor su estado de salud. De esta forma, la comparación de las tasas de incidencia de cardiopatía isquémica entre ambas cohortes estaría sesgada, ya que los sujetos con peor estado de salud y mayor probabilidad de desarrollar dicha enfermedad habrían rehusado participar en uno de los grupos de estudio.

## EXCLUSIONES PREASIGNACIÓN EN LOS ENSAYOS CLÍNICOS

En los ensayos clínicos, la repercusión de estos problemas tiene características que los diferencian de lo que ocurre en los estudios observacionales. Las llamadas *exclusiones preasignación* corresponden a los sujetos que, aunque inicialmente son candidatos a participar en el estudio, no son finalmente incluidos en él, ya sea porque no cumplen todos los criterios de selección o bien porque no consienten en participar. Si los criterios de selección son muy rigurosos y/o muchos sujetos rehúsan colaborar en el estudio, la muestra de pacientes realmente incluida diferirá mucho de la candidata, lo que limitará la generalización de los resultados. Además, algunos ensayos clínicos utilizan un período de preinclusión con la finalidad de excluir, antes del inicio del estudio propiamente dicho, a los sujetos que presentan efectos secundarios, problemas de cumplimiento, etc., lo que conformará una muestra en la que será más fácil observar el efecto que se busca, pero que será más difícilmente extrapolable.

En los ensayos clínicos, dado que estas exclusiones se han producido antes de la asignación de los sujetos a los grupos, no se vulnera el principio de comparabilidad y, por tanto, no se afecta la validez interna de los resultados. Sin embargo, a pesar de que no se analicen como parte del propio estudio, debe recogerse información sobre los motivos por los que no son incluidos, tanto si es porque no cumplen los criterios de selección como si es porque no desean participar en el estudio, o por cualquier otro motivo. Esta información será imprescindible para

evaluar la validez externa y su posible impacto sobre la extrapolación de los resultados.

## PÉRDIDAS

Se entiende por pérdidas aquellos casos que, habiendo iniciado el estudio, lo abandonan o se pierde el contacto con ellos, de forma que no es posible determinar el valor de la variable de respuesta. Habitualmente se producen por abandono del sujeto, que no desea continuar participando en el estudio, por pérdida durante el seguimiento o por la aparición de algún suceso concomitante durante éste que impide la valoración de la respuesta.

Según la naturaleza de la variable de respuesta que se utiliza puede realizarse algún esfuerzo suplementario para obtener información sobre ella en los sujetos que se han perdido. Por ejemplo, si se trata de la mortalidad, puede acudirse a los registros de mortalidad para obtener esta información. Sin embargo, en otras ocasiones, es imposible determinarla. Si se trata de la valoración del estado funcional de un paciente al mes de una intervención, o una toma de presión arterial al año de seguimiento, y no se ha determinado en ese momento, puede ser un dato que resulte imposible de obtener.

Si las pérdidas se producen aleatoriamente, es decir, por motivos no relacionados con lo que se está estudiando, no introducen un sesgo en la comparación de los resultados. Si su número es importante, pueden conducir a una disminución en la precisión de la estimación del efecto. En la tabla 23.1 se resumen las implicaciones que pueden tener estas situaciones en los ensayos clínicos aleatorios. Por otro lado, su posible impacto sobre los resultados depende de la frecuencia de presentación de la respuesta.

**Ejemplo 23.3.** En el cuadro 23.2 se presentan los datos de dos estudios hipotéticos de seguimiento de una cohorte de 100 pacientes, ambos con un 10% de pérdidas, pero que difieren en la frecuencia de aparición de la respuesta. En la situación A fallece el 60% de los pacientes, mientras que en la situación B, sólo el 3,3%. Puede observarse cómo la distorsión producida en la estimación del efecto es mayor en la situación B, cuando la respuesta se presenta con una frecuencia menor, en que la estimación del efecto puede ser hasta cuatro veces mayor en función de cómo se comporten las pérdidas.

**Ejemplo 23.4.** Imaginemos un estudio comparativo del efecto de dos fármacos, A y B, en el tratamiento de una enfermedad. La variable de respuesta es la

**Tabla 23.1 Exclusiones preasignación, pérdidas y retiradas en los ensayos clínicos aleatorios**

	Fase del estudio	Características	Motivos	Consecuencias
Exclusiones	Antes de la asignación	No son incluidos en el estudio	No cumplen los criterios de selección No dan su consentimiento para participar en el estudio	No producen sesgos Pueden limitar la generalización de los resultados
Pérdidas	Después de la asignación	No se ha determinado la variable de respuesta	Pérdidas y abandonos durante el seguimiento Sucedos concomitantes (otros procesos, etc.)	Si el motivo de la pérdida está relacionado con el pronóstico inicial o con algún efecto de la intervención, su exclusión del análisis puede producir un sesgo
Retiradas	Después de la asignación	Puede haberse determinado la variable de respuesta, pero se ha decidido deliberadamente su exclusión del estudio	Sujetos inelegibles (incluidos por error, etc.) Desviaciones del protocolo (fallos de cumplimiento, etc.) <i>Outliers</i>	Pueden producir sesgos importantes si la decisión de excluirlos del análisis está relacionada con el pronóstico inicial o con algún efecto de la intervención

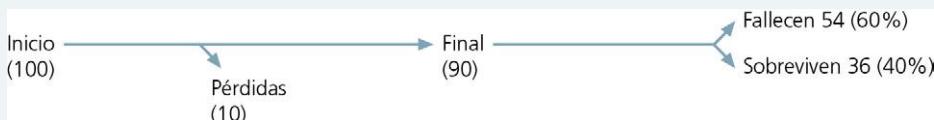
**Cuadro 23.2 Influencia de las pérdidas y retiradas del análisis sobre la estimación del efecto, en función de la frecuencia de presentación de la respuesta (ejemplo 23.3)**

Datos hipotéticos del seguimiento de una cohorte de 100 pacientes, durante el cual se han producido 10 pérdidas (10%). La variable de respuesta es la defunción del sujeto

Se plantean dos situaciones con diferentes valores observados de mortalidad

En cada una de ellas se estima el efecto potencial de las pérdidas en función de dos asunciones: el peor caso (las 10 pérdidas fallecen) y el mejor caso (las 10 pérdidas sobreviven) posibles

**Situación A.** Enfermedad de mal pronóstico: mortalidad 60%

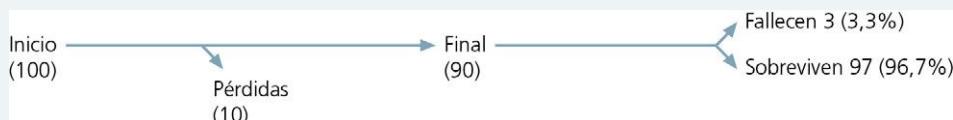


Peor caso posible: fallecen  $54 + 10 = 64$  (64%)

Mejor caso posible: fallecen  $54 + 0 = 54$  (54%)

Cociente entre ambas asunciones:  $64/54 = 1,2$

**Situación B.** Enfermedad de buen pronóstico: mortalidad 3,3%



Peor caso posible: fallecen  $3 + 10 = 13$  (13%)

Mejor caso posible: fallecen  $3 + 0 = 3$  (3%)

Cociente entre ambas asunciones:  $13/3 = 4,3$

mejoría en el estado funcional del paciente al año de seguimiento. Se diseña un ensayo clínico aleatorio, en el que se asignan 100 pacientes a cada uno de los grupos (tabla 23.2). Al año, el 90% de los pacientes que finalizan el estudio en cada uno de los grupos muestra mejoría, por lo que podría suponerse que su eficacia es idéntica. Sin embargo, al analizar los datos con detenimiento, puede observarse que en el grupo A se ha producido un 30% de pérdidas, mientras que en el grupo B sólo un 10%. Aun suponiendo que las pérdidas estén relacionadas con una menor eficacia de los tratamientos, y que ésta sea del 20% para ambos grupos, al tener en cuenta en el análisis todos los sujetos que han sido asignados, puede verse que el fármaco B es más eficaz (el 83 frente al 69%). Si el investigador se conforma con analizar a los sujetos que finalizan el estudio, puede llegar a conclusiones erróneas.

Este problema es muy similar al comentado para las no respuestas. La medición de la respuesta observada en los sujetos que han finalizado el estudio no es una buena estimación, ya que compromete la validez de la comparación. La mejor estrategia es, una vez más, diseñar el estudio de forma que se establezcan todos los mecanismos posibles para prevenir estas pérdidas y minimizar sus efectos. Y si, a pesar de todo, se producen, debe intentar obtenerse la máxima información de ellas para poder incluirlas en el análisis.

## MANEJO DE LAS PÉRDIDAS Y LAS NO RESPUESTAS

Una primera alternativa es considerar que las pérdidas o no respuestas se han producido al azar, y usar en el análisis sólo la información de los sujetos que han completado el estudio. El resultado de este análisis es equivalente a aplicar a las pérdidas la misma frecuencia de la respuesta observada entre

los que permanecen en el estudio. Un inconveniente que existe es la reducción de la precisión consiguiente a la disminución del número de sujetos, aunque, si se ha previsto la posibilidad de pérdidas o no respuestas al determinar el tamaño de la muestra, esta disminución puede ser contrarrestada. Sin embargo, la asunción de que las pérdidas o no respuestas se han producido al azar y de que no introducen ningún sesgo es habitualmente difícil de asegurar.

Una segunda alternativa es utilizar alguna técnica de ponderación de los resultados de la proporción observada de respuestas en diferentes estratos de la muestra. Las variables utilizadas para definir estos estratos deben ser predictoras del hecho de que los sujetos completen el estudio. Por ejemplo, ponderar los resultados por el inverso de la proporción de respuestas observadas en hombres y mujeres, y/o grupos de edad. Si se desea definir gran cantidad de estratos, el ajuste por este método es poco eficiente, ya que en algunas casillas habrá pocos sujetos y los factores de ponderación serán muy inestables. En esta situación es preferible utilizar una regresión logística para determinar las variables predictoras del hecho de completar el estudio, y utilizar esta información en la ponderación. El principal inconveniente de esta estrategia es la dificultad en el cálculo de las variancias. Por otro lado, la asunción que se realiza es que las pérdidas que se han producido han sido aleatorias, y el único hecho que se tiene en cuenta es que la proporción de sujetos que han completado el estudio difiere según determinadas características de ellos mismos.

Si se ha estudiado una submuestra de las pérdidas o las no respuestas para determinar en los estratos la variable de interés y evaluar la posible existencia de un sesgo, esta información puede utilizarse para calcular un factor de corrección para cada uno de los estratos e incluirlo en la ponderación. La asunción que se realiza es que la submuestra estudiada es una muestra aleatoria de todas las pérdidas o no respuestas que se han producido.

**Tabla 23.2 Resultados de un ensayo clínico hipotético comparativo de la eficacia de dos fármacos A y B, administrados a dos grupos de 100 pacientes. La variable de respuesta corresponde al cambio en el estado funcional al año de seguimiento. Se presentan los resultados por separado para los sujetos que han finalizado el estudio y para las pérdidas de seguimiento (ejemplo 23.4)**

Pacientes	Mejoría al año	
	Grupo A	Grupo B
Que han finalizado el seguimiento	63/70 (90%)	81/90 (90%)
Pérdidas	6/30 (20%)	2/10 (20%)
Total	69/100 (69%)	83/100 (83%)

Una tercera alternativa es asignar o imputar valores de la variable de respuesta u otras variables relevantes a las pérdidas o no respuestas. Existen diferentes opciones para realizar esta imputación de valores: extrapolar, para cada caso, los valores de la última visita realizada a las siguientes; interpolar los valores (si se dispone de datos anteriores y posteriores); calcular los valores según un modelo estadístico (se determina el valor de la respuesta de los sujetos en función de sus características basales, aplicando este modelo posteriormente a los sujetos con datos ausentes), y comparar los resultados de diferentes análisis, aplicando a los datos ausentes diferentes posibilidades de respuesta.

En general, las técnicas de imputación también asumen que las pérdidas y las no respuestas se han producido de forma aleatoria. Uno de los inconvenientes de estas técnicas es que reducen artificialmente la variabilidad de los valores, en especial cuando se asignan valores promedio en función de determinadas características, de forma que existirá una precisión artificialmente elevada, por lo que los intervalos de confianza deberán interpretarse con cautela.

Ninguna de las alternativas presentadas es ideal, y los estudios con gran cantidad de datos ausentes siempre son criticables. Sea cual fuere la opción utilizada, debe hacerse constar en el protocolo del estudio, y al publicar los resultados hay que explicar el procedimiento seguido y los valores asignados en lugar de los datos ausentes.

## RETIRADAS

Se entiende por retiradas aquellas personas incluidas en el estudio, pero que deliberadamente no son tenidas en cuenta en el análisis, a pesar de que en ellas se ha podido determinar la variable de respuesta. Este tipo de problemas se presenta, en general, en los estudios con seguimiento de sujetos, y ha sido ampliamente debatido en los ensayos clínicos, dadas las repercusiones que el análisis y la interpretación de estos estudios tienen sobre las decisiones terapéuticas. Las causas por las que puede plantearse la posible exclusión de un sujeto del análisis son muy variadas: personas incluidas por error o que no cumplen todos los criterios de selección, desviaciones del protocolo, sujetos que no han cumplido con la intervención prescrita, individuos que han recibido una intervención distinta de la asignada, etc.

En algunos casos, estas situaciones pueden implicar un porcentaje importante de sujetos, por lo que su eliminación del análisis puede sesgar los resultados. La comparabilidad inicial de los grupos

debe mantenerse a lo largo de todo el estudio. Si se excluye un número diferente de pacientes en cada uno de los grupos, puede alterarse dicha comparabilidad. Pero, aunque el número de retiradas sea el mismo en ambos grupos, sus motivos pueden divergir y afectar a un tipo diferente de individuos, por lo que también puede alterarse la comparabilidad. Por tanto, es fundamental que el diseño del estudio prevea las posibles fuentes de retiradas e intente evitarlas.

## Sujetos que no cumplen los criterios de selección

Puede ocurrir que los sujetos sean incluidos en el estudio y asignados a un grupo, pero que se considere a posteriori que no cumplen los criterios de inclusión. Se pueden distinguir dos situaciones: cuando se producen errores en la valoración de los criterios de inclusión (p. ej., al registrar la edad o el sexo), en alguna determinación de laboratorio o incluso en los criterios diagnósticos, debido a la necesidad de efectuar con rapidez la inclusión en algunos procesos agudos; o cuando debe instaurarse un tratamiento rápidamente y la confirmación de alguno de los criterios se obtiene posteriormente al inicio de la intervención.

**Ejemplo 23.5.** En un estudio se comparó la eficacia de la vasopresina y de la epinefrina en 324 pacientes que sobrevivieron a un paro cardíaco mientras estaban ingresados en un hospital. Cincuenta de estos pacientes fueron finalmente excluidos del estudio porque el paro ocurrió antes de la llegada al hospital y, en consecuencia, habían sido asignados erróneamente (Stiell et al, 2001).

En la primera situación, la exclusión del análisis de los pacientes que se han incluido por error puede sesgar los resultados, si, como puede suceder, el descubrimiento del error no se produce al azar. Por ejemplo, los pacientes que responden peor a un tratamiento o que presentan más efectos adversos son los que pueden llamar la atención de los investigadores, que, en consecuencia, pueden buscar con más ahínco errores en la evaluación de los criterios de inclusión y exclusión de estos pacientes.

**Ejemplo 23.6.** El estudio *The Anturane Reinfarction Trial* (1980) comparaba el tratamiento con sulfinpirazona frente a placebo en pacientes con infarto agudo de miocardio ([tabla 23.3](#)). De los 1.558 sujetos incluidos y asignados, 71 (38 del grupo tratamiento y 33 del grupo placebo) fueron considerados inelegibles a posteriori. Si se excluyen del análisis, puede

**Tabla 23.3 Resultados finales de mortalidad en el Anturane Reinfarction Trial (1980), que tienen en cuenta a los individuos que fueron eliminados del análisis por ser considerados inelegibles (ejemplo 23.6)**

	Mortalidad (%)		
	Grupo sulfpirazona	Grupo placebo	p*
Analizados	64/775 (8,3%)	85/783 (10,9%)	0,07
Asignados	74/813 (9,1%)	89/816 (10,9%)	0,20
Inelegibles**	10/38 (26,3%)	4/33 (12,1%)	0,13

\*Significación estadística.  
\*\*Pacientes considerados inelegibles posteriormente a su asignación a uno de los dos grupos.

apreciarse que el grupo de sulfpirazona presenta una menor mortalidad que el grupo de placebo, aunque no alcanza el nivel convencional de significación del 5%, pero se acerca bastante. En cambio, si se analiza a todos los pacientes asignados, las diferencias son menores y no significativas. Puede apreciarse, así mismo, que en el grupo de placebo los sujetos inelegibles presentan una mortalidad similar a la del resto de los sujetos (el 12,1 frente al 10,9%), mientras que en el grupo de tratamiento esta mortalidad es tres veces superior (el 26,3 frente al 8,3%).

Como regla general, los pacientes que no cumplen con los criterios de elegibilidad y que se hayan incluido por error deben permanecer en el estudio. Una excepción podría darse cuando el establecimiento de los criterios de inclusión y exclusión es difícil. En estas circunstancias, un comité ajeno al diseño y la ejecución del estudio, y que desconozca a qué grupo ha sido asignado cada paciente y los resultados relacionados con la variable de respuesta, podría valorar la información sobre los criterios de selección obtenida antes de la asignación.

En la segunda situación también conviene incluir a todos los sujetos en el análisis, incluso aunque después se demuestre que algunos no cumplen los criterios de inclusión, ya que los resultados reflejarán más fielmente lo que sucede en la práctica clínica, donde los médicos, en algunos procesos agudos, inician un tratamiento antes de disponer de toda la información diagnóstica.

**Ejemplo 23.7.** En un estudio que evaluó la efectividad de un tratamiento frente al virus de la influenza, el fármaco debía administrarse durante las primeras 48 horas después de haberse iniciado los síntomas, que, en muchas ocasiones son indistinguibles de los de otras infecciones (Treanor et al., 2000). En el protocolo del estudio se especificaba que sólo se incluiría a los pacientes que dieran un resultado

positivo a la infección en los cultivos o pruebas serológicas. Se asignaron 629 pacientes, de los cuales 255 (40%) no dieron un resultado positivo en las pruebas diagnósticas de infección por el virus de la influenza. Los investigadores informaron que en los 374 pacientes infectados, el tratamiento redujo la duración de la enfermedad en un 30%. Sin embargo, el análisis en el que se tenía en cuenta a los 629 pacientes inicialmente incluidos y asignados a los distintos grupos de estudio mostró una reducción del 22%, un resultado menos impactante. La exclusión retrospectiva de un número importante de pacientes en los que no se espera que el tratamiento produzca un beneficio, crea una falsa impresión de eficacia. De hecho, entre los que tuvieron la enfermedad, el tratamiento causó vómitos o náuseas en el 19% de los pacientes, y es de esperar que entre los 255 que no tenían la infección por el virus de la influenza, el porcentaje de pacientes con efectos adversos sería similar, pero sin el beneficio que produce el tratamiento.

En los estudios de cohortes también pueden presentarse este tipo de situaciones. Supongamos que se evalúa la asociación entre la exposición a un factor de riesgo y un determinado tipo de cáncer. Obviamente, ninguno de los sujetos debe presentar la enfermedad al inicio del estudio. Los casos en los que se detecte el cáncer a los pocos meses de iniciado suelen ser excluidos del análisis, al considerar que la enfermedad ya estaba presente desde el principio, aunque en una fase todavía indetectable.

### Cumplimiento insuficiente con la intervención

Si el objetivo de un estudio es comparar la eficacia de dos pautas terapéuticas, puede parecer lógico no considerar en el análisis a aquellos sujetos que no hayan cumplido adecuadamente con la intervención

asignada. Sin embargo, esta actitud puede comprometer la validez interna.

Los fallos de cumplimiento pueden deberse al sujeto, a su médico habitual o al propio investigador, y pueden estar motivados por la aparición de efectos secundarios de las intervenciones, pérdida de interés en permanecer en el estudio o aparición de cambios en la situación del sujeto, entre otras razones.

La validez de las comparaciones entre cumplidores y no cumplidores para determinar el efecto de un tratamiento asume que ambos grupos son similares en todas las características, excepto en el cumplimiento de la intervención. Sin embargo, no suele ser así.

**Ejemplo 23.8.** Al comparar los datos de los grupos en tratamiento con clofibrato y con placebo del The Coronary Drug Project (1980), puede apreciarse que ambos grupos muestran cifras similares de mortalidad a los 5 años. La técnica de medida utilizada fue el recuento de las cápsulas. Los pacientes fueron clasificados como buenos cumplidores si habían tomado al menos el 80% de las cápsulas prescritas. La mortalidad fue superior entre los no cumplidores, incluso en el grupo placebo (el 28,2 frente al 15,1%), lo que evidenció que se trataba de sujetos de características diferentes de los cumplidores.

Algunos autores proponen la eliminación de los no cumplidores del análisis, argumentando que, si no han seguido la intervención, no pueden beneficiarse de ella. Sin embargo, ya que el cumplimiento puede estar relacionado con la intervención, y de hecho puede considerarse un efecto de ella, no se recomienda esta actitud. Si los no cumplidores son más frecuentes en uno de los grupos, puede aparecer un sesgo en la comparación. Pero incluso si el porcentaje es similar en ambos grupos, puede ser debido a causas diferentes y afectar a un tipo diferente de sujetos.

En algunas situaciones, cuando el estudio busca establecer si un tratamiento es eficaz cuando se administra correctamente, puede utilizarse un período de preinclusión (*run-in phase*) que permita valorar el cumplimiento de los candidatos, y que sólo aquellos que sobrepasan los criterios definidos sean incluidos y asignados aleatoriamente, de forma que las características relacionadas con el cumplimiento sean distribuidas equitativamente en ambos grupos.

### Sujetos que reciben una intervención diferente de la asignada

Supongamos que se desea evaluar si una determinada estrategia de educación sanitaria grupal es eficaz

como medida complementaria en el control de los pacientes diabéticos. Para ello se diseña un ensayo clínico aleatorio, en el que los pacientes diabéticos del centro de salud que cumplen con determinados criterios de selección son asignados a dos grupos, uno de los cuales recibirá la nueva modalidad de educación sanitaria, y el otro, la individual habitual en la consulta. Puede ocurrir, y de hecho ocurre, que algunos pacientes tengan conocimiento de la realización de los grupos de educación, y acudan a ellos aunque hayan sido asignados al grupo de control. ¿Dónde se van a considerar estos pacientes? En realidad, han recibido la intervención, por lo que parece lógico incluirlos en el grupo de estudio. Sin embargo, habían sido asignados al grupo de control.

Si del grupo control se elimina a los pacientes que acuden a la educación grupal, probablemente los más motivados, se rompe la comparabilidad inicial establecida por la asignación aleatoria. Si se incluyen en el grupo de educación sanitaria grupal, aumenta la probabilidad de encontrar resultados favorables a ella. En el grupo individual, al eliminar a estos mismos pacientes, se disminuye el porcentaje de pacientes controlados.

La mejor estrategia para manejar estas situaciones es considerar a cada individuo en el grupo al que ha sido asignado, con independencia de la intervención que haya recibido (análisis por intención de tratar).

**Ejemplo 23.9.** El Heart Protection Study (2002) tuvo como objetivo evaluar si un tratamiento hipolipemiante podía reducir la morbilidad y la mortalidad cardiovascular en pacientes de alto riesgo, independientemente de sus cifras de colesterol. Se incluyó a 20.536 hombres y mujeres de 40-80 años de edad, con unas concentraciones plasmáticas de colesterol total de al menos 3,5 mmol/l, siempre que tuvieran en un plazo de 5 años un riesgo considerable de muerte a causa de unos antecedentes previos de enfermedad coronaria, enfermedad oclusiva de las arterias no coronarias, diabetes o hipertensión tratada, o una combinación de estos factores. Esta muestra fue distribuida aleatoriamente en dos grupos: uno de ellos recibió tratamiento hipolipemiante, y el otro, un placebo. Durante el seguimiento, el 17% de los pacientes asignados al grupo placebo recibieron un tratamiento hipolipemiante debido a que sus médicos consideraron que sus concentraciones de colesterol habían aumentado lo suficiente como para requerirlo. De todos modos, los investigadores analizaron los datos de estos pacientes como si hubieran recibido un placebo, ya que ésta era la intención inicial del tratamiento.

## Outliers

Un *outlier* es un valor aberrante o inusual en comparación con el resto. Su inclusión en el análisis puede distorsionar los resultados, sobre todo en los estudios con un número reducido de sujetos. Algunos autores recomiendan eliminar estos valores del análisis. Sin embargo, sólo deberían eliminarse si son claramente un error de medición, ya que, en caso contrario, su exclusión podría alterar los resultados, despreciando la información de alguna situación que puede ser muy interesante.

Es lícito argumentar que, si se utilizan pruebas estadísticas paramétricas, los *outliers* tienen una gran influencia. Como solución de compromiso en estos casos, puede procederse al análisis con estos valores y sin ellos. Si las conclusiones obtenidas en ambos análisis son diferentes, los resultados deberán ser interpretados con mucha precaución.

## MODALIDADES DE ANÁLISIS DE UN ENSAYO CLÍNICO ALEATORIO

Dado que las pérdidas y retiradas difícilmente se producen al azar, cualquier exclusión de sujetos del análisis puede alterar la comparabilidad de los grupos. Con el fin de preservarla, la única opción válida de análisis es evaluar a todos los pacientes incluidos en el estudio según el principio denominado *análisis según intención de tratar (intention-to-treat analysis)* o *según asignación aleatoria (as-randomized)*, según el cual se analiza a todos los pacientes como pertenecientes al grupo al que fueron inicialmente asignados, con independencia del tratamiento que hayan recibido realmente o de cualquier desviación del protocolo que se haya producido. Algunos autores matizan la definición de este principio, incluyendo en el análisis tan sólo a los sujetos que han recibido al menos una dosis del tratamiento.

Aunque pueda parecer extraño analizar en un grupo a sujetos que no han tomado el tratamiento asignado, o incluso que han recibido la intervención del otro grupo u otras medicaciones, si no se siguiese el principio de intención de tratar, se comprometería la comparabilidad de los grupos. De todas formas, si estas situaciones se producen en un gran número de sujetos, el estudio debería ser considerado como no evaluable.

**Ejemplo 23.10.** Supongamos un ensayo clínico que evalúa la eficacia de una intervención educativa grupal para reducir peso en pacientes obesos, en relación con la atención habitual que reciben estos

sujetos, y en el que se ha incluido a 180 pacientes, de los cuales 90 han sido asignados a cada uno de los grupos. Una vez iniciado el estudio, 10 pacientes del grupo control manifiestan su deseo de recibir la intervención grupal y acuden a las sesiones. Por otro lado, entre los individuos asignados a la intervención grupal, 20 no acuden a ninguna de las sesiones a pesar de que habían dado su consentimiento para participar en el estudio.

Una posibilidad sería excluir a estos sujetos del análisis, comparando los 70 sujetos del grupo de estudio que han recibido la intervención con los 80 del grupo control que recibieron la atención habitual. Sin embargo, hay que tener en cuenta que se excluyen, del grupo de intervención grupal, 20 sujetos que podrían considerarse los menos motivados para reducir peso, ya que rehusan acudir a los grupos, mientras que del grupo control se excluye a 10 sujetos que podrían considerarse muy motivados, ya que acuden a los grupos de educación de forma espontánea. Por tanto, la exclusión de estos 30 sujetos del análisis conduciría a una comparación sesgada entre los grupos.

Otra posibilidad sería analizar a los sujetos según la intervención que han recibido realmente, pero en este caso todavía se agravaría más el problema anterior.

La asignación aleatoria ha formado dos grupos de 90 sujetos, que son los que pueden considerarse comparables. Por ello, la única opción válida es analizar a los sujetos según el grupo al que fueron asignados.

Una alternativa diferente al análisis según intención de tratar es comparar sólo a los pacientes que han sido asignados a un grupo, han recibido el tratamiento correspondiente y han sido seguidos hasta el final del estudio. Esta estrategia se conoce como análisis de *casos válidos o según protocolo (per-protocol)*. Aunque esta alternativa pueda parecer más atractiva a primera vista, la exclusión de sujetos del análisis puede generar sesgos y conducir a conclusiones erróneas, en especial si estas situaciones se producen en las primeras fases del ensayo y en mayor número en uno de los grupos, ya que no aparecen por azar. Si, por ejemplo, algunos sujetos de un grupo no finalizan el estudio porque presentan acontecimientos adversos, su exclusión del análisis conducirá, quizás, a un sesgo favorable al tratamiento. Habitualmente se producen más retiradas por acontecimientos adversos entre los pacientes que reciben un tratamiento activo que entre los que reciben placebo. También puede conducir a sesgos la exclusión de sujetos que no finalizan el estudio porque se han curado, o porque han requerido un

**Tabla 23.4 Diferentes estrategias de análisis de los resultados de un ensayo clínico hipotético que compara la eficacia de la aspirina en la prevención cardiovascular frente a placebo (ejemplo 23.11)**

	Aspirina	Control
N.º de sujetos asignados	1.000	1.000
<b>Análisis por protocolo</b>		
Pacientes que cumplen con la intervención asignada	700	900
Pacientes que cumplen con la intervención asignada y presentan un evento cardiovascular	35 (5%)	90 (10%)
RAR	10% – 5% = 5%	
RRR	5%/10% = 0,50	
<b>Análisis por intención de tratar</b>		
Pacientes que no cumplen con la intervención asignada	300	100
Pacientes que no cumplen con la intervención asignada y presentan un evento cardiovascular	30 (10%)	8 (8%)
Incidencia de eventos cardiovasculares en los grupos formados por la asignación aleatoria	(35+30)/1.000 = 6,5%	(90+8)/1.000 = 9,8%
RAR	9,8% – 6,5% = 3,3%	
RRR	3,3%/9,8% = 0,33	

RAR: reducción absoluta de riesgo.

RRR: reducción relativa de riesgo.

tratamiento adicional por ineeficacia de la intervención recibida.

La estrategia por intención de tratar, además de mantener la lógica de la asignación aleatoria, hace más difícil encontrar diferencias entre los grupos, y protege frente a la conclusión errónea de que un nuevo tratamiento es mejor que el de comparación.

**Ejemplo 23.11.** En la tabla 23.4 se presenta el ejemplo hipotético de un ensayo que compara aspirina con placebo en la prevención cardiovascular, en el que se asignan 1.000 sujetos a cada grupo de tratamiento. En el grupo de intervención, 300 sujetos no toman la aspirina, de manera que presentan un resultado similar al del otro grupo. En el grupo control, 100 pacientes la toman de forma intermitente, por lo que presentan un resultado intermedio entre ambos grupos. La estrategia de análisis según protocolo conduce a una estimación de una reducción relativa de riesgo (RRR) del 50%. En cambio, si se utiliza la estrategia de análisis por intención de tratar, la estimación es del 33%. En el análisis según protocolo se han excluido 300 sujetos del grupo de intervención por mal cumplimiento del tratamiento. Sin embargo, es de esperar que en el grupo control haya aproximadamente 300 sujetos

de características similares, pero que no han sido detectados como malos cumplidores, ya que habían sido asignados a un grupo en que no debían recibir ninguna intervención, de manera que no son excluidos del análisis. Algo similar podría decirse de los 100 sujetos del grupo control que han tomado aspirina, por lo que las características de los sujetos que se analizan en cada grupo en el análisis por protocolo no son similares. Por ello, se considera que el resultado del análisis por intención de tratar es más convincente.

Así pues, dado que la estrategia por intención de tratar es la más conservadora, se considera la más adecuada para el análisis de los resultados de un ensayo clínico. Sin embargo, esto es cierto cuando el objetivo del estudio es evaluar si un nuevo tratamiento es mejor que el de referencia (ensayos de superioridad), pero no lo es cuando la hipótesis de estudio es de equivalencia o de no inferioridad. En estas situaciones, la estrategia de análisis por intención de tratar, al hacer más difícil encontrar diferencias entre los grupos, favorece llegar a conclusiones de equivalencia y, por tanto, no es la más conservadora. Ello no significa que en estos estudios sea más adecuada una estrategia de análisis por

protocolo, ya que no está claro el sesgo que puede introducir y no permite distinguir entre los efectos del tratamiento y el sesgo debido a las diferencias en el pronóstico de los sujetos, sino que probablemente lo más correcto es realizar ambos tipos de análisis:

si coinciden razonablemente en los resultados, la conclusión será más consistente, pero si difieren de forma sustancial, probablemente será difícil obtener una conclusión convincente.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet*. 2002;360:7-22.
- Stiell IG, Hébert PC, Wells GA, Vandemheen KL, Tang AS, Higginson IA, et al. Vasopressin versus epinephrine for inhospital cardiac arrest: a randomised controlled trial. *Lancet*. 2001;358:105-9.
- The Anturane Reinfarction Trial Research Group. Sulfur-pyrazone in the prevention of sudden death after myocardial infarction. *N Engl J Med*. 1980;302:250-4.
- 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;303:1038-41.
- Treanor JJ, Hayden FG, Vrooman PS, Barbarash R, Bettis R, Riff D, et al. Efficacy and safety of the oral neuraminidase inhibitor oseltamivir in treating acute influenza: a randomized controlled trial. US Oral Neuraminidase Study Group. *JAMA*. 2000;283:1016-24.

## BIBLIOGRAFÍA

- Begg CB. Ruminations on the intent-to-treat principle. *Control Clin Trials*. 2000;21(3):241-3.
- Clark TG, Altman DG, De Stavola BL. Quantification of the completeness of follow-up. *Lancet*. 2002;359:1309-10.
- Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ*. 2002;325:652-4.
- Fleming TR. Addressing missing data in clinical trials. *Ann Intern Med*. 2011;154:113-7.
- Goetghebeur E, Loey T. Beyond intention to treat. *Epidemiol Rev*. 2002;24:85-90.
- Grimes DA, Schulz KF. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet*. 2002;359:781-5.
- Gupta SK. Intention-to-treat concept: A review. *Perspect Clin Res*. 2011;2:109-12.
- Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9(1):48-55.
- Kreiger N, Nishri ED. The effect of non-response on estimation of relative risk in case-control study. *Am J Epidemiol*. 1997;146:194-9.
- Kruse RL, Alper BS, Reust C, Stevermer JJ, Shannon S, Williams RH. Intention-to-treat analysis: who is in? Who is out? *J Fam Pract*. 2002;51:969-71.
- Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trials*. 2000;21(3):167-89.
- Matts JP, Launer CA, Nelson ET, Miller C, Dian B, The Terry Beirn Community Programs for Clinical Research on AIDS. A graphical assessment of the potential impact of losses to follow-up on the validity of study results. *Stat Med*. 1997;16:1943-54.
- Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ*. 2001;165:1339-41.
- Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. *J Clin Epidemiol*. 2007;60:663-9.
- Wiens BL, Zhao W. The role of intention to treat in analysis of noninferiority studies. *Clin Trials*. 2007;4:286-91.

## Gestión del estudio

Muchos estudios fracasan por no haber sido gestionados de la forma adecuada. Una vez que los aspectos metodológicos del diseño están bien definidos, el reto es planificar la gestión del proyecto de manera que permita responder de forma efectiva a las necesidades del estudio y de los investigadores.

Dado que los recursos disponibles para la realización de la investigación son limitados, es fundamental garantizar que el estudio se lleva a cabo de la forma más eficiente posible.

Los aspectos clave de la gestión de un estudio no son diferentes de los de cualquier otro tipo de proyecto: definición clara del objetivo, formación del equipo de trabajo, definición del cronograma e identificación de los recursos necesarios y de las tareas que deben realizarse. La planificación de la gestión de la investigación debe contemplar los detalles para la ejecución y el seguimiento de todos los aspectos del estudio, tales como la gestión del equipo investigador y la comunicación entre sus miembros, la monitorización del reclutamiento de participantes, la gestión de la recogida de datos y del control de su calidad, la realización del análisis estadístico y la difusión y comunicación de los resultados.

Para que un estudio se ejecute de forma correcta, es necesario que todos los procedimientos se ajusten a lo planificado en el protocolo. Por ello, es conveniente elaborar un manual de procedimientos que contemple todas las situaciones posibles, de forma que los investigadores sepan qué hacer en todo momento y cómo actuar ante cualquier situación. En ocasiones, el manual puede estar incorporado en el mismo protocolo del estudio.

Por otro lado, antes de iniciar el estudio, debe haberse diseñado el cuaderno de recogida de datos (CRD) y formado y entrenado a las personas responsables de su cumplimentación. De la misma

forma, el resto de los documentos necesarios para la realización del estudio (hojas de consentimiento informado, de información al paciente, solicitudes de colaboración y participación, etc.) deben estar diseñados, y los investigadores, familiarizados con ellos.

Es aconsejable realizar una prueba piloto de los procedimientos y actividades previstas en el protocolo. Aunque el plan parezca perfectamente ensamblado en la teoría y los métodos que se van a utilizar estén especificados con detalle, puede ocurrir que, en la práctica, algo, o todo, funcione de diferente manera. Es preferible descubrir los problemas en un ensayo previo y realizar las modificaciones adecuadas antes de que el estudio se inicie.

### EL EQUIPO INVESTIGADOR

El equipo investigador está formado por el grupo de personas que se coordinan para llevar a cabo un determinado proyecto de investigación, y su composición viene definida por las necesidades del propio estudio. La elección de los miembros debe hacerse cuidadosamente, ya que es importante conseguir crear un grupo verdaderamente colaborativo de personas interesadas en desarrollar el proyecto en común.

El equipo debe incluir todos los perfiles profesionales necesarios para el desarrollo del proyecto, lo que a menudo implica a diferentes centros e instituciones. Los miembros deberían tener la formación, el entrenamiento y la experiencia adecuados para desarrollar sus tareas. Cuanto mayor sea su experiencia en investigación y su conocimiento del tema de estudio, mejor; sin embargo, suele ser interesante dar una oportunidad a investigadores

noveles para que se incorporen al equipo, aunque es conveniente mantener un equilibrio con los más experimentados.

El líder del proyecto suele denominarse investigador principal (IP), y es el máximo responsable. Debe ser un científico experimentado y su autoridad se basa más en la capacidad de liderazgo y la autoridad moral que en la imposición. En los estudios en que colaboran diferentes centros suele ser conveniente que exista un investigador coordinador en cada uno de ellos, el cual, además de ser responsable de la marcha del proyecto en su propio centro, debe participar en el grupo coordinador del estudio en su conjunto.

Es importante definir unas «reglas del juego» entre todos los miembros del equipo, de modo que se establezcan unas normas de funcionamiento claras, y que detallen y clarifiquen las tareas y responsabilidades de cada uno, para estimular sus habilidades personales y sin crear competencias desequilibradas, para fomentar su formación, establecer las posibles promociones (tesis, publicaciones, etc.) y los criterios y el orden de autoría de las publicaciones que se deriven, por ejemplo. La finalidad es conseguir un verdadero clima de colaboración y de sentimiento de pertenencia al equipo y de propiedad del proyecto, lo que suele ser un factor clave para el éxito del estudio. Aunque no es imprescindible que las personas que tan sólo participan de forma parcial o marginal en el estudio figuren como miembros del equipo, es importante prestarles mucha atención, y reconocer y motivar su participación, ya que pueden ser importantes para la buena marcha del estudio.

En el caso concreto de los ensayos clínicos se ha elaborado una «norma internacional de calidad científica y ética dirigida al diseño, realización, registro y redacción de informes de ensayos que implican la participación de seres humanos» (ICH E6), conocida como buena práctica clínica (BPC) o *good clinical practice* (GCP), con el fin de garantizar la credibilidad de los datos obtenidos en el estudio, así como que los sujetos de una investigación clínica mantengan íntegros sus derechos durante la misma. En esta guía, entre otros aspectos, se definen las obligaciones de los comités éticos de investigación, los investigadores (**cuadro 24.1**), el promotor y los monitores, así como algunos elementos relacionados con el manual del investigador y otros documentos esenciales.

Los investigadores deben sentir que forman parte de un equipo que intenta responder una pregunta importante, por lo que es recomendable asegurar un adecuado *feedback* de la información sobre la marcha del proyecto que contribuya a mantener su implicación en el mismo. Así mismo, es importante

#### Cuadro 24.1 Principales obligaciones del investigador en un ensayo clínico (ICH E6 de buena práctica clínica)

- Estar cualificado (titulación, formación y experiencia) para responsabilizarse de la realización correcta del ensayo clínico
- Conocer las propiedades de los medicamentos en investigación
- Conocer y ajustarse a la guía de BPC y a la normativa vigente
- Realizar el ensayo cumpliendo el protocolo
- Comprobar el uso correcto del producto en investigación en cada sujeto
- Disponer de los recursos apropiados para realizar y completar correctamente el ensayo dentro del tiempo convenido
- Dar al sujeto información oral y escrita concerniente al ensayo, y obtener su consentimiento informado firmado y fechado antes de su participación
- Responsabilizarse de las decisiones relacionadas con el estudio
- Proveer una asistencia médica adecuada a los sujetos del estudio
- Garantizar que todas las personas implicadas respeten la confidencialidad de cualquier información acerca de los sujetos
- Recoger, registrar y notificar los datos de forma correcta, tanto en el CRD como en la historia clínica
- Facilitar el acceso a los datos y documentos originales para la monitorización del ensayo y las auditorías e inspecciones de las agencias reguladoras
- Notificar inmediatamente los acontecimientos adversos graves o inesperados al promotor
- Informar regularmente al CEIC de la marcha del ensayo
- Corresponsabilizarse con el promotor de la elaboración del informe final del ensayo dando su acuerdo del mismo con su firma
- Guardar los documentos de consentimiento informado firmados, la lista de códigos de identificación de los pacientes, los CRD editados y el resto de los documentos del estudio durante 15 años tras la finalización del mismo
- Guardar la historia clínica de los sujetos y otros datos originales del estudio el máximo tiempo posible que permita la normativa

BPC: buena práctica clínica.

CRD: cuaderno de recogida de datos.

CEIC: Comité Ético de Investigación Clínica.

respetar sus opiniones, fomentar su participación y dar respuestas y soluciones rápidas a los problemas que se vayan planteando, lo que ayuda a aumentar la confianza en el equipo y evita que se sientan excluidos o marginados del mismo.

Los grandes estudios multicéntricos suelen requerir la figura de un gestor del estudio, encargado de dirigir el proyecto, vigilar el cumplimiento del calendario, gestionar el personal, etc. La importancia de esta figura es reconocida cada vez con más frecuencia, y se recomienda que se incorpore al equipo investigador en fases tempranas del diseño, ya que puede realizar aportaciones valiosas sobre aspectos prácticos de la ejecución del estudio. Aunque no es imprescindible que tenga amplios conocimientos del tema de estudio, debe estar familiarizado con el mundo de la investigación, y su perfil debe incluir la capacidad para organizar y motivar a otros, habilidades de comunicación y presentación, así como capacidad de gestión del presupuesto del proyecto y de liderar la coordinación y la ejecución del estudio.

## MANUAL DE PROCEDIMIENTOS

El manual de procedimientos puede considerarse una versión ampliada de la sección de métodos del protocolo del estudio, ya que define de forma operativa todos los métodos relacionados con el diseño y la organización del estudio, indicando de quién es responsabilidad cada tarea, dónde debe realizarse y en qué condiciones. Su función principal es estandarizar los procedimientos y disminuir la variabilidad en todas las fases del estudio, por lo que debe estar disponible para todos los investigadores en cualquier momento.

Al diseñar los aspectos prácticos del desarrollo del estudio, debe intentarse que supongan el menor trabajo posible para los investigadores y los participantes, por lo que es conveniente conocer detalladamente las características y el funcionamiento de los centros que participan. No debe olvidarse que los procedimientos complejos, el exceso de pruebas o visitas, o las dificultades para obtener determinados datos, pueden reducir el interés y la motivación de los investigadores y los participantes.

El manual de procedimientos ha de incluir: los métodos para la selección de la muestra y el reclutamiento de los sujetos; las instrucciones sobre la forma de ponerse en contacto con los participantes en el estudio; la definición y el procedimiento detallado para la medición de las variables e instrucciones para su codificación; los métodos de obtención y gestión de las muestras; los procedimientos para asegurar

la confidencialidad de los datos; las instrucciones claras de qué ha de hacerse en cada una de las visitas de seguimiento, cómo deben recaptarse las no respuestas o las pérdidas de seguimiento y qué hacer ante situaciones que planteen conflictos, entre otros. También ha de incluir todos los impresos que se van a utilizar en el estudio, como la hoja de recogida de datos, las cartas de presentación a los participantes o las hojas de consentimiento informado, así como los aspectos relacionados con la gestión de los datos y el control de calidad de éstos. En el cuadro 24.2 se presenta un ejemplo de índice de un manual de procedimientos.

Cuadro 24.2 Ejemplo de índice de un manual de procedimientos

1. Resumen del estudio:
  - Objetivos específicos del estudio
  - Características generales del diseño
2. Información general:
  - Investigadores y centros participantes; especificar el o los investigadores principales
  - Datos del promotor y fuentes de financiación
  - Centro coordinador (en estudios multicéntricos)
3. Selección de los sujetos:
  - Definición operativa de los criterios de selección
  - Plan de muestreo
  - Método de reclutamiento, incluir cómo y cuándo solicitar el consentimiento informado
  - Número de pacientes
4. Descripción de las intervenciones (en ensayos clínicos):
  - Método y procedimientos de aleatorización
  - Definición operativa de las intervenciones: descripción, dosis, pauta, etc.
  - Tratamientos concomitantes
  - Medicación de rescate
  - Evaluación del cumplimiento
5. Desarrollo del estudio:
  - Esquema de seguimiento
  - Definición de las variables
  - Procedimientos a realizar en cada visita, incluir los métodos de medición
  - Recaptación de las no respuestas y pérdidas de seguimiento
  - Procedimientos de seguimiento de los que no responden
6. Evaluación de la respuesta:
  - Técnicas de medida de la respuesta
  - Valoración de la respuesta
  - Acontecimientos adversos que dificultan la medida de la respuesta

(Continúa)

**Cuadro 24.2 Ejemplo de índice de un manual de procedimientos (cont.)**

7. Gestión y análisis de los datos:
  - Control de calidad de los datos
  - Entrada de datos
  - Archivo y copias de seguridad
  - Plan de análisis
8. Consideraciones prácticas:
  - Responsabilidades de los investigadores y otros profesionales implicados en el estudio
  - Entrenamiento de los procedimientos
  - Modificaciones del protocolo
  - Etiquetado y gestión de las muestras
  - Política de publicación
9. Bibliografía
10. Apéndices:
  - Hojas de recogida de datos, cuestionarios y formularios
  - Cartas de presentación y agradecimiento a los participantes
  - Hoja de información al paciente y consentimiento informado
  - Hoja de notificación de acontecimientos adversos
  - Valores de normalidad de los parámetros de laboratorio

y el procesamiento de los datos. Debe ser fácil de manejar, evitando el uso de letra menuda o apretada y de hojas de pequeño formato, lo que disminuirá el número de errores. Cada vez es más frecuente el uso de CRD electrónicos, e incluso de un procedimiento de entrada telemática de los datos.

Es conveniente encabezar el CRD con el título del proyecto y registrar la fecha de recogida de los datos para evitar posibles confusiones con otros estudios llevados a cabo en el mismo centro o por el mismo profesional.

Es preferible utilizar un CRD para cada individuo (o unidad de estudio), identificándolo en todas las hojas, habitualmente mediante un número secuencial. Ello evita confusiones y permite verificar posteriormente algunos datos y localizar al paciente si es necesario su seguimiento. Sin embargo, no debe olvidarse la confidencialidad de la información contenida en estos documentos. Puede ser útil mantener un fichero separado en el que consten sus datos de identificación.

El CRD debe seguir la secuencia en que los datos serán recogidos. Después de la identificación del estudio y del sujeto, debe recogerse la información que permite comprobar que el candidato cumple con los criterios de inclusión y exclusión exigidos, las exploraciones que delimiten el estado inicial del sujeto y, posteriormente, el resto de las variables en orden cronológico. Deben agruparse los ítems por visitas, lo que evita tener que retroceder varias hojas para anotar algún dato durante una visita.

Las páginas deben ser espaciosas, con contenidos distribuidos ampliamente y bien alineados. Deben resaltarse los encabezamientos y los aspectos más importantes con caracteres tipográficos de diverso tamaño o subrayándolos. Incluso puede ser aconsejable utilizar algún tipo de ilustración. Hay que evitar el uso de abreviaturas no habituales.

Si existen varias páginas, éstas deberían estar identificadas con un sistema de acceso rápido mediante indicaciones del número de visita, o bien utilizar hojas o indicadores de diferentes colores.

Es conveniente que el CRD especifique todas las instrucciones y codificaciones necesarias para su cumplimentación. Además de facilitar la recogida de datos, puede ser un complemento importante del protocolo, al incluir una lista de los criterios de inclusión y exclusión, o de las exploraciones que deben realizarse en cada visita.

## CUADERNO DE RECOGIDA DE DATOS

El CRD es el soporte documental específicamente diseñado para recoger los datos del estudio, de forma que toda la información relevante de cada uno de los sujetos quede registrada y disponible.

Debe estar preparado antes del inicio de la recogida de datos, y haber sido elaborado con la suficiente atención. Sin embargo, no se diseña hasta que el plan de estudio no está prácticamente completado, ya que debe contemplar las definiciones y escalas de medida de las variables implicadas, así como la pauta de seguimiento (número de visitas, exploraciones que hay que realizar, periodicidad de las mediciones, etc.). Antes del inicio de la recogida de datos es conveniente discutir el diseño del documento con todos aquellos que deban cumplimentarlo, procesarlo o analizarlo.

### Formato

El CRD debe ser de diseño tan claro y atractivo como sea posible, y estar dirigido a facilitar la recogida

### Precodificación

Debe facilitarse el procesamiento de los datos. Lo ideal es un documento precodificado en el que el

investigador deba escribir lo menos posible. No debe obligarse a las personas que recogen los datos a hacer cálculos antes de codificarlos, sino que es preferible entrar los datos brutos y hacer los cálculos directamente en la fase de análisis. Puede ser conveniente reservar algún espacio para comentarios y observaciones que permita registrar información complementaria relevante.

Pueden identificarse diferentes tipos de datos precodificados:

- *Datos alfanuméricos* que deben escribirse completamente (p. ej., nombre y apellidos).
- *Datos dicotómicos*. Preferiblemente debe utilizarse una codificación numérica (1: sí / 2: no), evitando el dígito 0 para no confundirlo con los datos ausentes. Puede ser útil prever un código para la respuesta «no se sabe», que idealmente debe ser el mismo para todas las variables. Los códigos de las respuestas deben ser consistentes en todas las preguntas.
- *Respuestas múltiples*. Deberían ser mutuamente excluyentes y contemplar todas las situaciones posibles. Si no son excluyentes, es preferible desagregar las respuestas y considerarlas como variables aisladas.
- *Datos numéricos*. Es aconsejable que las centenas, las decenas y las unidades ocupen siempre la misma posición, y ajustar las cifras por la derecha. Debe tenerse un cuidado especial en la señalización de los puntos decimales. Es preferible sobreestimar que infraestimar el número de dígitos necesarios. No es conveniente categorizar variables continuas que pueden recogerse directamente, por ejemplo, la edad.
- *Fechas*. Las fechas suelen introducirse con el formato día/mes/año.
- *Datos no precodificados* cuando existen demasiadas respuestas posibles; por ejemplo, otros procesos concomitantes, tratamientos previos, etc.

Es aconsejable utilizar el mismo código para las no respuestas en todas las variables. En cualquier caso, deben evitarse los espacios en blanco que puedan indicar tanto la ausencia de respuesta como que ésta no ha sido valorada, no ha sido considerada relevante o se ha olvidado.

Si bien el CRD ha de ser autoexplicativo y contener las instrucciones necesarias para su cumplimentación, no debe sustituir al manual de procedimientos que explique detalladamente cómo debe recogerse toda la información, e incluya las definiciones de las variables, las instrucciones para los investigadores y el circuito que debe seguir la información recogida.

## FORMACIÓN DEL PERSONAL

La calidad de los datos depende en gran medida de las personas que los recogen, de su experiencia, su conocimiento del protocolo, su motivación por el estudio y, muy especialmente, de su meticulosidad. Las personas más cualificadas para recoger los datos no son necesariamente las más competentes desde el punto de vista profesional. Muchos investigadores prestan poca atención a la recogida de datos porque la consideran tediosa y menos estimulante desde el punto de vista intelectual que las fases de diseño y análisis.

Debe recogerse información sobre todas las variables consideradas de interés, de forma homogénea por todos los investigadores, para todos los sujetos y durante todo el estudio. Para ello, además de la definición clara de las variables, de la fuente de datos, del instrumento de medida y de las codificaciones utilizadas, debe realizarse una formación adecuada de las personas que participarán en la recogida.

La primera estrategia es elegir un número reducido de personas adecuadas y entrenarlas correctamente. Las sesiones de formación y entrenamiento de los investigadores y personas que participen en la recogida de la información para promover su homogeneidad son cruciales, sobre todo en estudios de gran tamaño. Debe discutirse con ellos la importancia de una recogida exhaustiva y exacta de los datos. De la misma forma, deben calibrarse y validarse los instrumentos de medida.

**Ejemplo 24.1.** Supongamos que se desea realizar un estudio sobre pacientes con demencia. Dada la dificultad que entraña el diagnóstico de la enfermedad, existe mucha variabilidad entre los profesionales. Para mejorar las habilidades y homogeneizar los criterios diagnósticos, se decide utilizar una estrategia de formación y entrenamiento en la que se incluyen actores que simulan ser pacientes con demencia. Se graban las entrevistas y se debaten con los investigadores para corregir los errores y homogeneizar los criterios.

**Ejemplo 24.2.** En un estudio sobre el tratamiento de la hipertensión arterial se instruyó a los profesionales que recogían los datos en las técnicas para localizar la arteria braquial, aplicar el manguito de presión, insuflar y desinsuflar el manguito, y reconocer los sonidos que indican la presión arterial diastólica. Además, se les realizó una prueba teórica escrita sobre los contenidos del manual de procedimientos y una prueba práctica de la toma de la presión arterial (Siegel et al, 1987).

Si los datos que se han de recoger son muy objetivos (p. ej., los resultados de una analítica), la formación debe ir dirigida principalmente al registro cuidadoso y detallado de los datos. Si el estudio requiere la recogida de información mediante encuestas o entrevistas, quienes las hayan de realizar deberán someterse a una formación y entrenamiento específicos sobre cómo contactar con el sujeto, cómo formular las preguntas y cómo clarificar y registrar las respuestas.

### Contactar con el participante

En el contacto inicial, la persona que recoge los datos debe presentarse, identificarse, indicar la institución que promueve la investigación y explicar sus objetivos. El entrevistador debe estar preparado para responder a preguntas del tipo: ¿cómo me seleccionaron?, ¿quién le dio mi nombre?, ¿para qué sirve todo esto?, ¿qué más tendré que hacer?, ¿cómo sé que estos datos son confidenciales?, ¿por qué quiere saber todo esto?, ¿qué van a hacer con todas estas respuestas?

### Formular las preguntas

Para aumentar la homogeneidad de la recogida de datos, es importante que quienes la llevan a cabo dispongan de instrucciones precisas y detalladas sobre cómo formular las preguntas y clarificar las respuestas (cuadro 24.3). La estandarización en los métodos de entrevista es esencial. Cada pregunta se

ha de formular tal y como está en el cuestionario, sin interpretaciones ni aclaraciones adicionales, ya que pequeñas variaciones pueden comportar cambios en el significado de la pregunta.

Cuando un participante no entienda una pregunta, el encuestador deberá repetirla nuevamente en su totalidad. Si el entrevistado todavía muestra sus dudas, el encuestador no debe intentar explicársela (ya que podría indirectamente inducir la respuesta), sino que puede usar alguna frase neutra del estilo: «lo que signifique para usted», para clarificar la pregunta. Si aun así el participante no está seguro de responder, la pregunta se deja sin respuesta y se anota en el cuestionario la razón de esta dificultad.

### Clarificar las respuestas

En ocasiones, el entrevistador usa frases para clarificar una respuesta cuando ésta ha sido incompleta o irrelevante. Estas frases sirven para motivar al participante a responder de forma más completa y ayudarle a centrar una cuestión. Estos objetivos se han de cumplir sin sesgar las respuestas.

**Ejemplo 24.3.** Un participante podría responder «lo normal» a la cuestión: «¿cuántos cigarrillos fuma usted diariamente?». Esta respuesta puede considerarse inaceptable, ya que «normal» puede significar un cigarrillo, cinco, diez, un paquete o dos, según la persona. En este caso, es conveniente que el entrevistador formule una segunda pregunta: «¿qué entiende usted por normal?». Esta pregunta es mucho más neutra que la siguiente: «¿qué entiende usted por normal, un paquete?». Hay que entrenar a los encuestadores para que formulen estas preguntas aclaratorias de la forma más neutra posible.

Los encuestadores deben estar entrenados en el uso de frases u otros métodos útiles para clarificar la información:

- *Repetir la pregunta.* Una respuesta imprecisa puede ser consecuencia de que el encuestado no ha entendido la pregunta o no ha tenido el tiempo suficiente para pensarla. En esta situación, el encuestador debe repetir la pregunta con una introducción como la siguiente: «no estoy seguro de haberle comprendido. Si me lo permite, le formularé nuevamente la pregunta...». Hay que tratar de clarificar las respuestas del tipo «no sé». A veces puede ser útil decirle al participante una frase como la siguiente: «no hay respuestas buenas o malas para estas preguntas; simplemente responda lo que usted crea que es más correcto».

Cuadro 24.3 Recomendaciones para la recogida de datos mediante un cuestionario

- Leer las preguntas exactamente como están formuladas en el cuestionario
- Leer despacio cada una de las preguntas
- Usar una entonación correcta
- Preguntar las cuestiones en el orden en que están en el cuestionario
- Preguntar sólo las cuestiones que afectan a los participantes
- Repetir en su totalidad una pregunta que haya sido malinterpretada por un participante
- Usar sólo las frases permitidas para clarificar respuestas
- Leer las frases que enlazan bloques de preguntas como están impresas en el cuestionario
- No dar explicaciones por propia cuenta, a menos que estén escritas en el cuestionario

- *Pausa de espera* para indicar que el encuestador está esperando más información, que no le basta con la que le ha proporcionado hasta ese momento.
- *Repetir la respuesta* para estimular al participante a proporcionar más detalles.
- *Preguntas o comentarios neutros*. Se utilizan con frecuencia para clarificar (*¿qué es lo que quiere decir exactamente?*, *¿qué es lo que usted entiende por...?*), conseguir más especificidad (*¿podría ser más específico sobre esto?*, *¿podría darme una fecha más concreta?*) u obtener una información más completa (*¿podría decirme alguna cosa más sobre esto?*, *¿hay alguna otra razón por la que usted piense de esta manera?*).

### Registrar las respuestas

Como norma general, un encuestador debe registrar de forma exacta y completa cada una de las respuestas. Si se deja en blanco hay que anotar la razón (rechazo del entrevistado, falta de comprensión, etc.).

Las opiniones del encuestado deben aceptarse con toda naturalidad, y el entrevistador nunca manifestará sorpresa, rechazo ni aprobación. Antes de registrar una respuesta, el encuestador debe estar seguro de que la ha comprendido suficientemente. Las respuestas se han de registrar mientras se efectúa la entrevista, utilizando las mismas palabras que el encuestado. Una vez finalizada la entrevista, debe revisarse el CRD para comprobar que no se ha omitido ninguna pregunta y que no falta ninguna información.

Es frecuente que al responder una pregunta el encuestado facilite accidentalmente información sobre preguntas posteriores. El problema es que, si el encuestador formula una pregunta cuya respuesta ya ha sido facilitada con anterioridad, el encuestado puede molestarse al pensar que no se le presta la suficiente atención.

**Ejemplo 24.4.** Supongamos que al realizar la siguiente pregunta: «*¿alguna vez su médico le ha encontrado o le ha dicho que tuviera el colesterol alto en la sangre?*», el encuestado responde: «*sí, y también el azúcar*». Si después se pregunta sobre los antecedentes de diabetes, no sería correcto por parte del encuestador preguntarlo directamente. Ante este tipo de preguntas es conveniente usar alguna frase aclaratoria como: «*antes me dijo algo sobre esta cuestión, pero no he tenido tiempo de anotarlo, ¿le importaría repetirlo?*», y, a continuación, formular la pregunta tal y como está en el cuestionario.

## SEGUIMIENTO DEL ESTUDIO

Un aspecto crucial de cualquier investigación es garantizar que todos los aspectos relacionados con el mismo se han realizado con el máximo de rigor y calidad, por lo que es importante desarrollar procedimientos normalizados de trabajo para las diferentes tareas que deben llevarse a cabo. En caso de un ensayo clínico, deben seguirse escrupulosamente las directrices establecidas en la norma de BPC (ICH E6).

Un estudio, especialmente si es de gran envergadura, necesita un sistema y unos procedimientos adecuados de monitorización de todos los aspectos del día a día del estudio: reclutamiento de los sujetos, control del equipamiento y material, gestión y revisión de los datos, etc.

Deben elaborarse informes periódicos de la marcha del proyecto, que deben remitirse a los investigadores y ser discutidos con todas las personas que colaboran en el estudio con la finalidad de llevar a cabo las acciones oportunas para corregir los problemas detectados, informando de ello al equipo coordinador del estudio.

Es especialmente importante realizar un seguimiento del ritmo de reclutamiento de los sujetos. En caso de que sea más lento de lo esperado en algunos centros, deben analizarse sus posibles causas y soluciones. Si es más rápido de lo previsto, debe estudiarse si se están cumpliendo los criterios de selección de los sujetos.

**Ejemplo 24.5.** Una revisión de la Colaboración Cochrane sobre los incentivos y la desmotivación para la participación en los ensayos clínicos encontró 11 estudios observacionales que relacionaban las tasas de reclutamiento de los estudios con diferentes factores (Rendell et al, 2007). En concreto, sugerían que el reclutamiento era más elevado cuando el investigador:

- Estaba interesado en los aspectos de la medicina basada en la evidencia.
- Participaba en un grupo académico.
- Disponía de personal adicional que le ayudaba en el proceso de reclutamiento.
- Pensaba que los pacientes deberían estar interesados en el estudio.
- Se sentía cómodo al explicar el estudio.

A pesar de que estos aspectos pueden ser de ayuda a la hora de seleccionar los colaboradores en una investigación, los autores de la revisión concluyen que la evidencia disponible es de poca calidad y que se necesita más y mejor investigación.

En algunas ocasiones, especialmente en los ensayos clínicos, se cuenta con profesionales que realizan el seguimiento detallado de la marcha del estudio, conocidos como monitores, cuya finalidad fundamental es verificar que el investigador sigue el protocolo y realiza las funciones específicas del ensayo, comprobar la fiabilidad e integridad de las anotaciones hechas en el CRD, contrastándolas con los documentos originales y verificando la exactitud de los datos, comprobar que el almacenamiento, la distribución, la devolución y la documentación de los medicamentos son seguros y adecuados, y que se guardan los documentos esenciales, etc. El monitor debe presentar informes de monitorización después de cada visita.

## CONTROL DE CALIDAD DE LOS DATOS

Ya que las conclusiones del estudio se basarán en el análisis de los datos recogidos, hay que prever mecanismos para asegurar su calidad. Las estrategias para mejorarla pueden aplicarse en distintas fases del proyecto:

- *En el diseño*, elaborando un manual detallado y preciso de todos los procedimientos y diseñando un instrumento adecuado para la recogida de los datos.
- *En las fases previas al inicio de la recogida*, formando a las personas responsables.
- *Durante la recogida*, supervisando el proceso y controlando la exhaustividad y la calidad de los datos.
- *Durante su manejo*, asegurando su correcta codificación y proceso informático.

### Control de calidad de los datos durante la recogida

Durante la realización del estudio, la verificación de los datos permitirá detectar inexactitudes, errores u omisiones, datos incorrectamente recogidos o resultados dudosos o improbables. Este proceso debe realizarse de forma simultánea a la recogida de datos para permitir la recuperación o la rectificación de la información.

Antes de dar por finalizada cualquier visita o entrevista con un participante, debe comprobarse que se han recogido todos los datos. No debería archivarse ningún CRD que no estuviera completo. Si se requiere la utilización de instrumental, éste debe revisarse periódicamente, aunque haya sido

calibrado inicialmente. Lo mismo ocurre si se almacena material que puede deteriorarse con el tiempo (p. ej., tiras reactivas), y prever las condiciones necesarias de almacenaje.

A veces puede ser de utilidad volver a entrevistar a una muestra de los participantes para detectar datos ficticios, ítems poco fiables, encuestadores que se equivocan sistemáticamente en las mismas preguntas, etc. También es útil comparar los datos recogidos por los distintos observadores para ver si difieren sistemáticamente o valorar la evolución de las medidas de variabilidad de la variable de respuesta a medida que avanza el estudio para detectar un posible deterioro en la fiabilidad de las mediciones.

Todos estos mecanismos de control de la calidad de los datos son más valiosos cuando existe una estrecha relación con los profesionales que los recogen, intentando solucionar los problemas a medida que se detectan. Deben realizarse reuniones para reentrenar a las personas que recogen los datos, discutir los problemas que hayan aparecido y mantener la motivación.

### Supervisión de los responsables de la recogida de datos

Además de la formación inicial, los responsables del estudio deben seguir la evolución de las personas que recogen los datos de la siguiente forma:

- *Supervisar los CRD* y corregir los errores que se detecten.
- *Buscar posibles variaciones de un encuestador a otro*. Si la población es similar y se observan resultados diferentes entre los entrevistadores puede ocurrir que se esté introduciendo un error sistemático.
- *Impartir durante el estudio alguna sesión de recuerdo complementada con entrevistas supervisadas*. El principal inconveniente de esta estrategia es su coste. Sin embargo, permite que los encuestadores refresquen conceptos y se den cuenta de la importancia de seguir con fidelidad las instrucciones. El simple hecho de saber que los responsables del estudio controlan la calidad de las encuestas hace que los entrevistadores se esfuerzen en seguir fielmente el manual de instrucciones.

### Procesamiento de los datos

Antes de la introducción de los datos, deben revisarse los CRD para detectar errores u omisiones de información o problemas de legibilidad y, si aparecen, intentar corregirlos o verificarlos.

También existen mecanismos para verificar los datos a medida que se introducen. Uno de los más

útiles es la doble entrada, que consiste en introducir dos veces cada uno de los registros, con el fin de comparar ambas versiones y detectar inconsistencias. Otro es utilizar filtros que impiden introducir valores fuera del rango previamente definido o dejar sin informar determinados campos.

A medida que se introducen los datos, deben revisarse periódicamente las distribuciones de las variables buscando valores extremos o ilógicos y corrigiéndolos. Los datos registrados incorrectamente pueden llegar a ser indetectables, por lo que lo mejor es evitar que aparezcan. Hay que evaluar la coherencia de los datos. Por ejemplo, no pueden anotarse pulsos periféricos simétricos en extremidades inferiores en una persona con antecedentes de amputación, ni cambios de gran magnitud en el peso de una persona entre dos visitas próximas. También deben corroborarse los valores improbables obtenidos en exploraciones o exámenes complementarios.

## PRUEBA PILOTO

El número de pruebas piloto y sus características dependerán de la naturaleza y la envergadura del estudio. Su finalidad es poner a prueba los métodos y procedimientos del estudio antes de su inicio: métodos de reclutamiento de sujetos, de medición de variables, de gestión de datos, etc. Pueden proporcionar estimaciones sobre el número de sujetos disponibles y dispuestos a participar, la eficiencia de diferentes métodos de reclutamiento, datos para el cálculo del tamaño de la muestra, así como evaluar la factibilidad de algún método de obtención de datos (p. ej., la existencia de determinada información en la historia clínica), evaluar la reproducibilidad de algunas mediciones, la validez y la fiabilidad de un instrumento de medida o un cuestionario, etc. En el cuadro 24.4 se enumeran los pasos que seguir en la prueba piloto y el desarrollo de la versión final del CRD.

Antes de iniciar el estudio puede ser conveniente poner a prueba todos los procedimientos de reclutamiento y medición y, en general, los aspectos organizativos y logísticos del estudio, en un ensayo general en el que se incluyan individuos que cumplen los criterios de selección. Aunque en el papel puede parecer un protocolo que no tenga problemas de ejecución, en la práctica pueden aparecer multitud de dificultades que deberán subsanarse antes de iniciar el estudio.

### Cuadro 24.4 Pasos a seguir en la prueba piloto y el desarrollo de la versión final del cuaderno de recogida de datos (CRD)

1. Solicitar a los colegas la revisión del CRD
2. Evaluar el CRD en una muestra de conveniencia
3. Revisar las instrucciones para el uso del CRD
4. Entrenar a los responsables de la recogida de datos para la prueba piloto
5. Realizar una prueba piloto en una muestra de 30-50 sujetos de características similares a la población de estudio
6. Recoger las opiniones de los responsables de la recogida de datos y participantes, preferiblemente por escrito
7. Revisar las preguntas que presentan alguna dificultad
8. Realizar una nueva prueba piloto y revisar el CRD
9. Preparar las instrucciones revisadas y entrenar a los responsables de la recogida de datos
10. Monitorizar las primeras fases de administración del CRD

## DIFUSIÓN Y COMUNICACIÓN DE LA INVESTIGACIÓN

Un estudio no es útil si sus resultados no son difundidos y tenidos en cuenta en la práctica clínica. La difusión puede realizarse de muy diversas maneras: artículos en revistas impresas, artículos *online*, páginas web, presentaciones a congresos, sesiones clínicas, etc. En capítulos posteriores se abordan los temas relacionados con la comunicación científica.

Un aspecto importante de la gestión de un proyecto es planificar cómo se difundirán los resultados y cómo se compartirá el crédito por el estudio, definiendo las publicaciones que se prepararán, quién y en qué orden firmará cada una de ellas, etc.

El reconocimiento adecuado de todas las personas que han contribuido suficientemente al estudio es vital, lo que puede ser un elemento adicional de incentivación y motivación. De forma similar, debe asegurarse que todas las personas que han realizado méritos suficientes constan como miembros del grupo colaborativo en todas las ocasiones en que se comunique el estudio.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- Rendell JM, Merritt RD, Geddes JR. A Cochrane review on incentives and disincentives to participation by clinicians in trials. *Cochrane Database Syst Rev.* 2007;18. MR000021.
- Siegel D, Kuller L, Lazarus NB, Black D, Feigal D, Hughes G, et al. Predictors of cardiovascular events and mortality in the Systolic Hypertension in the Elderly Program pilot project. *Am J Epidemiol.* 1987;126(3): 385-99.

## BIBLIOGRAFÍA

- Baer A, Zon R, Devine S, Lyss AP. The Clinical Research Team. *J Oncol Pract.* 2011;7:188-92.
- Caloto T. Multicentre Project for Tuberculosis Research Study Group. Quality control and data-handling in multicentre studies: the case of the Multicentre Project for Tuberculosis Research. *BMC Med Res Methodol.* 2001;1(1):14.
- Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, Knight R, et al. Recruitment to randomised trials: strategies for trial enrollment and participation study. The STEPS study. *Health Technol Assess.* 2007; 11:iii, ix-105.
- Fairhurst K, Dowrick C. Problems with recruitment in a randomised controlled trial of counselling in general practice: causes and implications.
- J Health Serv Res Policy. 1996;1:77-80.
- Farrell B, Kenyon S, Shakur H. Managing clinical trials. *Trials.* 2010;11:78.
- Gassman JJ, Owen WW, Kuntz TE, Martin JP, Amoroso WP. Data quality assurance, monitoring and reporting. *Control Clin Trials.* 1995;16(2 Supl):104S-36S.
- ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice (E6). Adopted by CPMP, July 1996, issued as Note for a Guidance on Good Clinical Practice (CPMP/ICH/135/95). Disponible en: <http://www.ich.org>.
- Irving SY, Curley MA. Challenges to conducting multicenter clinical research: ten points to consider. *AACN Adv Crit Care.* 2008;19:164-9.
- Knatterud GL. Management and conduct of randomized controlled trials. *Epidemiol Rev.* 2002;24:12-25.
- Moss AJ, Francis CW, Ryan D. Collaborative clinical trials. *N Engl J Med.* 2011;364:789-91.
- Rebagliato M, Ruiz I, Arranz M. Metodología de investigación en epidemiología. Madrid: Díaz de Santos; 1996.
- Sprague S, Matta JM, Bhandari M. Anterior Total Hip Arthroplasty Collaborative (ATHAC) Investigators. Multicenter collaboration in observational research: improving generalizability and efficiency. *J Bone Joint Surg Am.* 2009;91(Suppl3):80-6.
- Whitney CW, Lind BK, Wahl PW. Quality assurance and quality control in longitudinal studies. *Epidemiol Rev.* 1998;20:71-80.

## Solicitud de ayuda para la financiación

Todos los proyectos o protocolos de investigación deberían pasar algún proceso de evaluación externa que avale su interés y calidad científica. Muchas instituciones recomiendan, e incluso obligan, a su evaluación por parte de un Comité Ético de Investigación Clínica (CEIC), que valore además el cumplimiento de los principios éticos de la investigación en seres humanos, así como la protección y la confidencialidad de los datos que se van a recoger. De hecho, cada vez más revistas establecen como requisito imprescindible para la publicación de un artículo de investigación la aprobación previa por un CEIC, y no solamente para los ensayos clínicos con medicamentos. Además, en muchos casos, los investigadores solicitan algún tipo de ayuda para poder llevar a cabo el estudio, lo que implica el envío del protocolo para su evaluación por parte de alguna agencia de financiación.

Un proyecto tendrá una alta probabilidad de llevarse a la práctica con éxito si el equipo investigador tiene los conocimientos y la experiencia suficientes sobre el tema de estudio para formular una pregunta clara y precisa, si el diseño del estudio es el adecuado para responderla, si los métodos de selección de la población y de recogida de la información garantizan que la respuesta será válida, si el tamaño de la muestra y el tiempo de seguimiento (si lo hay) son suficientes, si se dispone de los medios técnicos y presupuestarios necesarios para llevar a cabo las actividades y si se respetan los principios éticos y la normativa legal vigente. Por tanto, cualquier evaluador del proyecto prestará una especial atención a estos aspectos.

### **PREPARACIÓN DE UNA SOLICITUD DE AYUDA PARA LA FINANCIACIÓN**

En muchas ocasiones se requiere financiación externa para poder llevar a cabo un proyecto de investigación. En estos casos, además del protocolo de estudio, hay que llenar una solicitud de ayuda para la financiación, dirigida a las agencias o instituciones que las conceden. Las convocatorias de ayudas para la financiación de proyectos de investigación incluyen formularios estandarizados para presentar tanto los aspectos del diseño contemplados en el protocolo como la información adicional administrativa y de financiación que es exigida por la agencia que concede las ayudas. La función de estos formularios es conseguir que los distintos grupos de investigación presenten la información de forma similar, y facilitar así la labor de los evaluadores.

Esta solicitud de ayuda no sustituye el protocolo de estudio. De hecho, su enfoque es algo diferente, ya que las agencias financieras no sólo evalúan la calidad metodológica del proyecto, sino que realizan también una evaluación estratégica de la adecuación de la propuesta a las prioridades de la política sanitaria o de la propia agencia, y una evaluación económica que se basa en la adecuación del presupuesto al proyecto evaluado y en la disponibilidad presupuestaria de la agencia. Por tanto, con la propuesta se trata de convencer a los evaluadores, en poco espacio, de la bondad del proyecto, de su

pertinencia, de que los resultados esperables tendrán aplicabilidad práctica, de la capacidad del equipo investigador para desarrollarlo y de que el presupuesto está acorde con la envergadura del estudio.

Para elaborar una solicitud de ayuda de financiación no es necesario tener un protocolo de estudio totalmente finalizado que contenga todos los detalles del estudio. No se trata de extenderse en detalles metodológicos y definiciones operativas, sino de exponer la estructura general del diseño con el fin de que la agencia pueda evaluar si los aspectos metodológicos son correctos y permitirán obtener una respuesta válida y fiable.

En el cuadro 25.1 se presentan los principales elementos que contempla la solicitud de una ayuda del Instituto de Salud Carlos III (ISCIII) a un proyecto de investigación en salud. Los contenidos metodológicos de la solicitud se han de exponer de forma escueta, y deben quedar especialmente claros el objetivo y la pregunta o preguntas concretas que se quieren responder, así como la arquitectura básica del estudio. Los apartados de experiencia del equipo investigador y de medios necesarios y disponibles están destinados a evaluar la idoneidad y la capacidad del equipo para poder llevar a cabo el estudio. La investigación requiere una dedicación sustancial, sin la cual es improbable la obtención de logros importantes, por lo que es necesario que los equipos tengan el personal y la dedicación suficientes. De la misma forma, la mayoría de las agencias no ven favorablemente que un mismo investigador participe en múltiples proyectos de forma simultánea. La ayuda solicitada debe justificarse con detalle, teniendo en cuenta que sólo podrán solicitarse conceptos directamente relacionados con el proyecto y sólo, excepcionalmente, personal técnico y auxiliar.

Conjuntamente con la memoria del proyecto debe enviarse el compromiso del organismo que gestiona el centro o los centros donde se realizará el estudio, así como la composición del equipo investigador y un currículum normalizado de cada uno de sus miembros. La mayoría de las agencias exigen que algún miembro del equipo investigador ostente la función de investigador principal, que será quien dirija los esfuerzos y asuma el compromiso con la agencia financiadora. Además, algunas agencias solicitan que la propuesta se acompañe de un informe del CEIC que avale el proyecto.

Hay que tener presente que cada agencia financiadora tiene su propio modelo de solicitud, a veces con exigencias específicas. Por tanto, lo primero es decidir a qué agencia se solicitará la ayuda, y a continuación obtener un modelo de solicitud y de las instrucciones para su cumplimentación, que deben seguirse estrictamente. Si no se tiene mucha

### Cuadro 25.1 Apartados de la memoria de solicitud de ayuda de proyectos de investigación en salud (Instituto de Salud Carlos III)

#### • Identificación del proyecto:

- Título
- Investigador/a principal
- Tipo de proyecto: individual, coordinado, multicéntrico

#### • Resumen:

- Objetivos y metodología del proyecto

#### • Antecedentes y estado actual del tema:

- Finalidad del proyecto, antecedentes y estado actual de los conocimientos científico-técnicos, grupos nacionales o internacionales que trabajan en la línea específica del proyecto o en líneas afines

#### • Bibliografía más relevante

#### • Hipótesis y objetivos

#### • Proyectos coordinados:

- Objetivos globales, necesidad y valor añadido de la coordinación
- Objetivos específicos de cada subproyecto
- Interacción entre los distintos objetivos, actividades y subproyectos
- Mecanismos de coordinación previstos

#### • Metodología:

- Diseño, sujetos de estudio, variables, recogida y análisis de datos y limitaciones del estudio

#### • Plan de trabajo:

- Etapas de desarrollo y distribución de las tareas de todo el equipo investigador
- Lugar/centro de realización del proyecto
- Cronograma: actividades/tareas y personas involucradas

#### • Experiencia del equipo investigador sobre el tema

#### • Plan de difusión:

- Relevancia en cuanto a su impacto clínico, asistencial y/o desarrollo tecnológico
- Relevancia en cuanto a su impacto bibliométrico

#### • Medios disponibles para la realización del proyecto

#### • Justificación detallada de las partidas presupuestarias solicitadas

#### • Presupuesto solicitado

#### • Anexos

tenerse en cuenta que el proyecto será evaluado por revisores expertos en investigación, pero que pueden no tener un conocimiento muy detallado del área concreta de interés, por lo que es fundamental presentar el proyecto de forma clara y concisa, con un lenguaje transparente y de manera que contenga toda la información necesaria para comprenderlo sin esperar que el revisor acuda a fuentes externas o adicionales. Antes de escribir la versión final es conveniente que la propuesta sea revisada por colegas que estén familiarizados con este proceso. Es mejor tener críticas inteligentes y detalladas antes de presentar la propuesta que tener que aceptar que ésta sea rechazada por problemas de forma, por una presentación poco persuasiva o porque no prevé soluciones a problemas potenciales.

## EVALUACIÓN DE LAS SOLICITUDES DE AYUDAS PARA LA FINANCIACIÓN

El propósito de la evaluación de las solicitudes es seleccionar los proyectos a los que se va a conceder la ayuda. El procedimiento que se sigue habitualmente tiene dos fases. La primera es una evaluación científico-técnica cuya finalidad es asegurar que se cumplen los requisitos éticos, la calidad científica del proyecto y su viabilidad, es decir, la capacidad para llevar a cabo el estudio, tratando de evitar que se pongan en marcha iniciativas con una alta probabilidad de fracaso o abandono. Para ello se valoran no solamente los aspectos metodológicos del proyecto presentado, sino también otros, como su adecuación a las prioridades de la política sanitaria o los aspectos económicos, y también aspectos relacionados con el propio equipo de investigación, como su estructura y composición, su nivel de experiencia y dedicación, o la consecución de los objetivos en las ayudas solicitadas previamente. En la segunda fase, en función de las disponibilidades presupuestarias, se seleccionan los proyectos teniendo en cuenta, además del resultado de la evaluación científico-técnica previa, aspectos como la adecuación de los proyectos a los objetivos y prioridades de la convocatoria, su interés, su potencial de transferencia real de los resultados a la práctica clínica, la cooperación entre grupos investigadores estable, la internacionalización de la actividad investigadora, etc.

La evaluación de un proyecto incluye tanto la revisión de cada uno de los apartados de la propuesta como los aspectos formales de la solicitud. El principal requerimiento formal es la presencia de toda la documentación solicitada. La falta de algún

documento impedirá que un proyecto llegue a la etapa de revisión científica. Aunque pueda parecer poco razonable denegar una ayuda a un proyecto por defectos de forma, parece existir una relación directa entre la calidad científica del proyecto y su adecuada presentación en tiempo y forma.

Actualmente, el principal mecanismo de control de la calidad de un proyecto de investigación es su revisión por expertos independientes, proceso conocido como revisión por pares (*peer-review*), que juzgan su pertinencia y si cumple los requisitos mínimos de calidad para alcanzar sus objetivos. Este sistema de revisión no está exento de errores, sobre todo cuando se ha de revisar un número elevado de proyectos en un tiempo limitado. Para minimizar estos errores, conviene que cada proyecto sea revisado por al menos dos expertos de forma independiente y, en caso de discrepancias, resolverlas mediante el consenso entre revisores o recurrir a un tercer revisor. Así mismo, el investigador que presenta el proyecto debe tener derecho de réplica si considera que su trabajo no se ha valorado de forma adecuada.

Los evaluadores de los proyectos han de ser imparciales, evaluar el contenido científico y sus méritos o defectos intrínsecos desde una postura de máxima neutralidad. Tampoco deben tener ningún conflicto de intereses con los proyectos que se evalúan. Así, por ejemplo, un evaluador no debería revisar un proyecto presentado por él mismo o por compañeros de trabajo, proyectos en los que tenga intereses económicos o financieros, o que sean presentados por grupos competidores en su misma línea de investigación. La documentación que se entrega a los evaluadores para revisión es confidencial y no debe ser divulgada ni utilizada en beneficio propio, por ejemplo, para anticiparse en el desarrollo de una idea.

Por otro lado, el evaluador debe ser competente científica y técnicamente, y estudiar a fondo el proyecto con el fin de formular el dictamen con criterio. No debe tomar decisiones rápidas ni caprichosas, ni ofender a los autores. Las críticas al proyecto deben efectuarse con corrección y profesionalidad, sin valoraciones peyorativas de la capacidad de los investigadores.

A continuación se comentan los principales aspectos que tienen en cuenta los evaluadores en cada uno de los apartados de la solicitud.

### Antecedentes y estado actual del tema

Este apartado permite a los evaluadores determinar el grado de conocimiento por parte de los investigadores acerca de la cuestión de estudio, así como la pertinencia del proyecto.

Esta sección debe redactarse de forma escueta y precisa, y ha de estar orientada hacia las hipótesis de trabajo. Los investigadores deben presentar los conocimientos principales existentes en el campo de investigación y qué preguntas quedan por responder. La descripción de generalidades y la falta de concreción de los problemas que quedan por resolver pueden indicar que los investigadores no están muy familiarizados con la materia, aunque ello no signifique necesariamente una falta de competencia en la práctica clínica.

### Bibliografía más relevante

Esta sección sirve de apoyo a la anterior; en ella deben constar los artículos más relevantes, siempre relacionados con la hipótesis de trabajo, que pueden ayudar a llevar a cabo el proyecto.

Debe procurarse que las referencias bibliográficas sean actuales, preferentemente de los últimos 5 años. Si se solicita que se comenten, debe especificarse por qué se consideran relevantes para la investigación.

### Objetivos concretos

Los objetivos de un proyecto son la declaración explícita de la pregunta o las preguntas que se pretende responder con el estudio, por lo que los expertos evaluarán los distintos apartados de la propuesta en relación con ellos.

En general, el número de objetivos de un proyecto debe ser limitado, y hay que diferenciar el motivo principal de los secundarios.

Aunque es difícil evaluar los objetivos de forma aislada del resto del proyecto, su definición clara y precisa es tan importante que la mayoría de los evaluadores le conceden la misma importancia que a otros apartados de la propuesta, como la metodología o la justificación. Por ello, los objetivos deben ser evaluables, factibles y relevantes tanto desde el punto de vista científico como social y sanitario. Si no se cumplen estos criterios, los evaluadores informarán negativamente la solicitud.

Los principales criterios para evaluar la relevancia científica de una propuesta son su originalidad, su capacidad para solucionar un problema de forma precisa y la posibilidad de producir avances significativos en el tema de estudio. La relevancia sociosanitaria depende de la repercusión socioeconómica del problema estudiado, ya sea por su frecuencia, por la carga que supone para el segmento de población afectado o por el impacto previsible que se derive del logro del objetivo del proyecto.

### Metodología y plan de trabajo

Este apartado es una síntesis de los elementos metodológicos más importantes detallados en el

protocolo de estudio. Han de presentarse de forma precisa todos los pasos que hay que seguir para alcanzar los objetivos propuestos, de forma que el evaluador no deba realizar ninguna asunción sobre lo que pretenden hacer los investigadores.

Esta información suele presentarse estructurada en epígrafes para facilitar su comprensión, habitualmente como población de estudio, diseño, descripción de las intervenciones (si procede), determinaciones y medidas del estudio, mecanismos de control de calidad, estudios piloto, descripción del plan de análisis estadístico, así como las dificultades y limitaciones del estudio y las etapas para su desarrollo. El evaluador debe juzgar si la metodología es adecuada para el problema de estudio y si permite obtener una respuesta válida a la pregunta planteada.

También debe detallarse el plan de trabajo, definiendo las diferentes etapas previstas para el desarrollo del proyecto, así como la distribución de las actividades y tareas entre los miembros del equipo investigador, con su correspondiente cronograma.

### Experiencia del equipo investigador sobre el tema

La mejor garantía de que un proyecto puede realizarse adecuadamente es que los investigadores hayan completado proyectos similares con anterioridad. Debe evaluarse, por tanto, la capacidad del equipo investigador en relación con la envergadura del proyecto. Para ello, el evaluador debe tener en cuenta no solamente los logros obtenidos por el equipo investigador en los años previos, sino también aspectos del currículum profesional de los investigadores.

El investigador principal debe acreditar su participación anterior en otros proyectos de investigación. Los evaluadores juzgan esta capacidad sobre todo por las ayudas concedidas previamente y por las publicaciones realizadas por los investigadores en los últimos 5 años, prestando atención no sólo a la calidad de los estudios y a su grado de contribución, sino también a la coherencia de las publicaciones en relación con el tema de estudio. Por ello, deben hacerse constar los artículos originales, ya que las monografías y los libros, editoriales o artículos de revisión pueden ser importantes desde un punto de vista académico, pero poco relevantes desde el de la experiencia investigadora.

El número de personas que integran el equipo investigador debe corresponderse con la envergadura del proyecto, y las responsabilidades de cada uno de sus miembros deben quedar claramente establecidas. La composición del equipo debe reflejar las necesidades del proyecto, con expertos en las técnicas que se van a utilizar durante su realización.

Un equipo consolidado y con experiencia aumenta las probabilidades de completar de forma satisfactoria el proyecto. La presencia de investigadores con experiencia es imprescindible cuando el investigador principal carece de ella.

### Plan de difusión

Se trata de valorar la relevancia del proyecto, juzgando los beneficios que los resultados previsibles aportan tanto a la sociedad como al propio sistema sanitario, para lo cual es importante tener en cuenta la capacidad para generalizar y extrapolar sus conclusiones y su aplicabilidad práctica.

Por otro lado, tienen importancia socioeconómica aquellos proyectos con elevada probabilidad de resultar en patentes, aplicaciones tecnológicas, programas informáticos o productos potencialmente comercializables.

En este apartado también debe valorarse el potencial impacto bibliométrico de las publicaciones que puedan derivarse del proyecto, ya que se utiliza como un indicador de la productividad científica.

### Medios disponibles

El evaluador juzga si la disponibilidad de medios materiales (instalaciones, instrumentos, equipa-

miento, técnicas, etc.) de los investigadores, tanto de los existentes como de los solicitados, es suficiente para realizar el proyecto.

### Justificación detallada de la ayuda solicitada

Los evaluadores han de revisar el presupuesto solicitado, juzgar si se ajusta a la envergadura del proyecto y si existe algún valor añadido que pueda aportar a la institución, como la consolidación de un grupo investigador y el fomento de relaciones entre instituciones, entre otros.

El personal contratado con cargo al proyecto se refiere a profesionales que se están formando como investigadores, y no, por ejemplo, a personal auxiliar para recoger datos o realizar tareas administrativas. Por tanto, debe tratarse de proyectos de cierta envergadura que justifiquen la necesidad de uno o más becarios, solicitados por equipos que tengan experiencia en la formación de personal investigador. Difícilmente se considera que un proyecto de corta duración permita la formación de un investigador.

En cuanto a las instalaciones y los equipos, el material fungible, los viajes, las dietas y otros gastos, sólo deberán solicitarse si son relevantes para el desarrollo del estudio.

## BIBLIOGRAFÍA

- Bellavista J, Guardiola E, Méndez A, Bordons M. Evaluación de la investigación. Colección «Cuadernos metodológicos», n.º 23. Madrid: Centro de Investigaciones Sociológicas; 1997.
- Devine EB. The art of obtaining grants. Am J Health Syst Pharm.. 2009;66:580-7.
- Espinosa de los Monteros J, Díaz V, Toribio MA, Rodríguez Farré E, Larraga V, Conde J, et al. La investigación biomédica en España (I). Evaluación del Fondo de Investigación Sanitaria (FIS) a través de los proyectos de investigación financiados en el periodo 1988-1995 a instituciones sanitarias asistenciales (hospitales). Med Clin (Barc). 1999;112:182-97.
- Espinosa de los Monteros J, Díaz V, Toribio MA, Rodríguez Farré E, Larraga V, Conde J, et al. La investigación biomédica en España (II). Evaluación del Fondo de Investigación Sanitaria (FIS) a través de los proyectos de investigación financiados en el periodo 1988-1995 a centros de investigación, facultades y escuelas. Med Clin (Barc). 1999;112:225-35.
- Guallar E, Conde J, De la Cal MA, Martín-Moreno JM. Grupo de evaluación de la actividad del Fondo de Investigaciones Sanitarias entre 1988 y 1995. Guía para la evaluación de proyectos de investigación en ciencias de la salud. Med Clin (Barc). 1997;108:460-71.
- Guillén-Grima F, Annan JW, Álvarez JM, Gómez JM, Ontoso EA. How to apply for research grants in allergology. Allergol Immunopathol (Madr). 2009;37:146-54.
- Inouye SK, Fiellin DA. An evidence-based guide to writing grant proposals for clinical research. Ann Intern Med. 2005;142:274-82.
- Prieto Carles C, Gómez-Gerique J, Gutiérrez Millet V, Veiga de Cabo J, Sanz Martul E, Mendoza Hernández JL. Análisis del proceso evaluador de los proyectos de investigación en el Fondo de Investigación Sanitaria. Med Clin (Barc). 2000;115(11):418-22.
- Trilla A. El proceso formal de evaluación de proyectos de investigación en el Fondo de Investigación Sanitaria. Enferm Clin. 2007;17:32-6.
- Van Ekelenburg H. The art of writing good research proposals. Sci Prog. 2010;93 (Pt 4):429-42.

# Parte

4

## Interpretación de resultados

- |   |  |
|---|--|
| <b>26</b> Interpretación de resultados <b>249</b> | <b>30</b> Confusión y modificación del efecto <b>280</b>         |
| <b>27</b> Papel de la estadística <b>253</b>      | <b>31</b> Inferencia causal <b>292</b>                           |
| <b>28</b> Potencia de un estudio <b>265</b>       | <b>32</b> Aplicabilidad práctica<br>de los resultados <b>298</b> |
| <b>29</b> Sesgos <b>272</b>                       |  |

Página deliberadamente en blanco

## Interpretación de resultados

Al finalizar un estudio se obtiene un resultado principal, ya sea en forma de estimación de un parámetro, de una diferencia entre grupos o de una asociación entre variables, con el que se pretende responder a la pregunta planteada en el objetivo.

Para interpretar este resultado, en primer lugar debe tenerse en cuenta la variabilidad aleatoria, ya sea cuantificando la precisión con que se ha realizado la estimación mediante el cálculo de su intervalo de confianza (IC), o bien estableciendo si la diferencia o la asociación es estadísticamente significativa. En segundo lugar, debe evaluarse si este resultado es válido descartando posibles sesgos que puedan haber conducido a una estimación distorsionada, evaluando si se han controlado los potenciales factores de confusión y rechazando explicaciones alternativas. Y, en tercer lugar, si se considera que el resultado es válido, debe evaluarse su utilidad práctica teniendo en cuenta su magnitud, su importancia y relevancia clínicas, así como su impacto potencial y su aplicabilidad en otros entornos y grupos de sujetos, etc.

Los responsables de la investigación se hallan en una posición privilegiada para cumplir esta función, ya que conocen mejor que nadie las limitaciones del estudio y de los datos recogidos. Tienen la responsabilidad de revisarlos críticamente y con honestidad, teniendo en cuenta los aspectos del diseño que puedan influir en su interpretación y evitar obtener conclusiones precipitadas o insuficientemente justificadas sobre objetivos no contemplados en el protocolo.

En este capítulo se introducen los principales aspectos que es necesario considerar para interpretar adecuadamente los resultados de un estudio, que serán desarrollados ampliamente en los capítulos siguientes.

Tras el análisis de los resultados de un estudio que compara dos o más grupos, pueden plantearse dos

situaciones generales: o bien se ha encontrado una diferencia estadísticamente significativa entre los distintos grupos respecto a la variable de respuesta (estudio positivo), o bien la diferencia observada no es estadísticamente significativa (estudio negativo) (fig. 26.1).

### PAPEL DE LA ESTADÍSTICA

En los estudios cuya finalidad es estimar un parámetro poblacional a partir de una muestra de sujetos (p. ej., estimar la prevalencia de una enfermedad en una comunidad determinada), la estadística permite cuantificar la precisión con que se ha realizado dicha estimación mediante el cálculo de un intervalo de valores (el IC) entre cuyos límites se tiene una determinada confianza de que esté incluido el verdadero valor poblacional del parámetro.

**Ejemplo 26.1.** Para estimar la prevalencia de hipertensión arterial (HTA) en una comunidad, se estudia una muestra representativa de sujetos y se encuentra que el 20% de ellos presenta HTA. Los autores indican que la prevalencia estimada es del 20%, con un IC del 95% cuyos límites son del 18 y el 22%. Este resultado significa que se tiene un 95% de confianza de que el intervalo entre el 18 y el 22% incluya el verdadero valor de la prevalencia de HTA en la comunidad de referencia.

En los estudios en los que se comparan dos o más grupos, el resultado que se obtiene es una diferencia de incidencias o de medias, o bien una medida relativa de asociación como la *odds ratio* o el riesgo relativo. En este caso, la estadística también permite cuantificar mediante un IC la precisión con que se

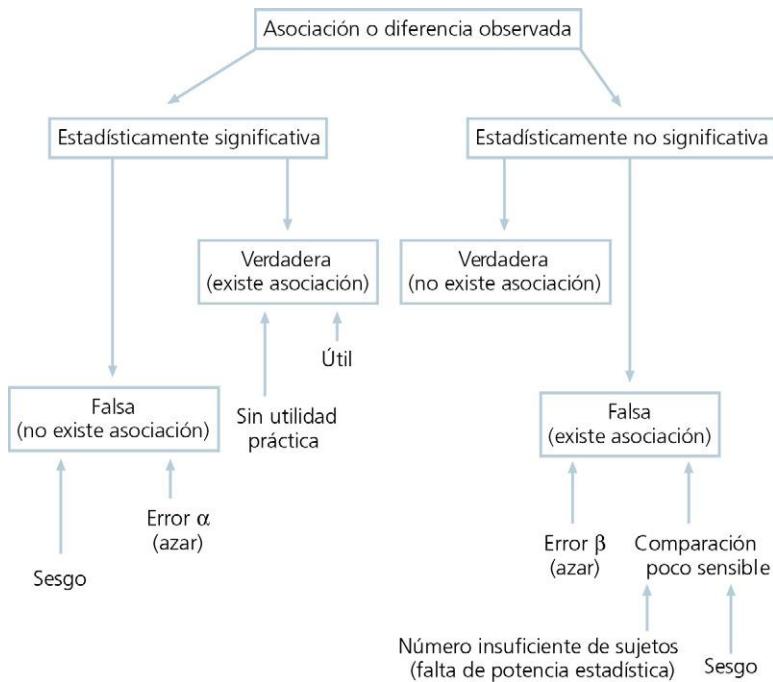


Figura 26.1 Aspectos a considerar en la interpretación de los resultados de un estudio.

ha estimado esta diferencia o asociación, así como determinar si este resultado es estadísticamente significativo mediante la utilización de una prueba estadística de contraste de hipótesis.

Un resultado estadísticamente significativo significa que, si en realidad no existe ninguna diferencia, es poco probable que el resultado observado en el estudio se deba simplemente a la variabilidad aleatoria, lo que conduce a la conclusión de que probablemente refleja una diferencia real entre los grupos. Sin embargo, no debe interpretarse que necesariamente el factor de estudio sea el responsable de esta diferencia, ya que pueden existir otras explicaciones alternativas que deban descartarse, como sesgos o factores de confusión no controlados.

Por otro lado, un resultado estadísticamente no significativo quiere decir que, si en realidad no existe ninguna diferencia, es relativamente probable que el resultado observado en el estudio pueda haberse producido simplemente por variabilidad aleatoria. No debe interpretarse como indicativo de que no existe una diferencia en la realidad, sino tan sólo de que los resultados del estudio no permiten descartar esta posibilidad. En esta situación, es importante determinar la potencia estadística del estudio, es decir, su capacidad para haber detectado una determinada diferencia si existiera en la realidad.

**Ejemplo 26.2.** Supongamos un estudio que compara un grupo de sujetos con hipercolesterolemia que reciben un tratamiento hipolipemiante frente a otro grupo con el mismo problema de salud pero que no recibe ninguna intervención. Se observa que el primer grupo ha reducido sus cifras de colesterolemia en 10 mg/dl más que el grupo de comparación. Tras aplicar una prueba estadística de contraste de hipótesis, se concluye que este resultado es estadísticamente significativo ( $p=0,03$ ), lo cual quiere decir que, en el supuesto de que el tratamiento no fuera eficaz, existiría un 3% de probabilidades de producirse una diferencia como la observada o mayor, simplemente por azar. Dado que puede considerarse que esta probabilidad es pequeña, se concluye que el primer grupo de sujetos ha presentado una mayor reducción de la colesterolemia que el grupo de comparación. Sin embargo, no puede asegurarse que el tratamiento sea el responsable de esta diferencia, sino que deberá evaluarse si el diseño del estudio era adecuado, si los grupos eran comparables, si la estrategia de análisis ha sido la correcta, si pueden haber existido otros factores que hayan podido interferir, etc. Además, si se llega a la conclusión de que el tratamiento es la única explicación plausible de la diferencia observada, de ello tampoco puede deducirse que sea una buena

alternativa terapéutica de la hipercolesterolemia, sino que deberá evaluarse la relevancia clínica de un descenso promedio de 10 mg/dl, los efectos secundarios del tratamiento, su coste, otras ventajas e inconvenientes de su administración, la existencia de otras alternativas terapéuticas, etc.

## SESGOS Y FACTORES DE CONFUSIÓN

Para evaluar si el resultado observado proporciona una respuesta válida a la pregunta formulada en el objetivo, deben descartarse errores en el diseño del estudio que puedan comprometer su validez interna. De hecho, se trata de comprobar hasta qué punto puede concluirse que el factor de estudio causa el efecto o la asociación observados.

Los sesgos son errores sistemáticos introducidos en un estudio tanto en el proceso de selección de los sujetos como en el de recogida de la información, que conducen a estimaciones erróneas del resultado. No sólo hay que evaluar la posible existencia de estos errores, sino también, en caso de que existan, su tendencia a sobreestimar o infraestimar el resultado, y valorar su magnitud y sus implicaciones prácticas.

**Ejemplo 26.3.** En un estudio de casos y controles cuyo objetivo era evaluar la hipótesis de que la endometriosis predispone a la infertilidad, se revisaron las historias clínicas de 100 mujeres a las que se había practicado una laparoscopia a causa de su infertilidad (Strathy et al., 1982). En 21 de ellas se diagnosticó endometriosis. Como grupo control se eligieron 200 mujeres a las que se había practicado una laparoscopia para una ligadura de trompas. En este grupo control sólo se diagnosticó endometriosis en cuatro mujeres. El valor de la *odds ratio* fue de 13. Se puede argumentar que la información sobre la endometriosis no es comparable en los casos y en los controles, ya que no es lo mismo buscar este proceso cuando la laparoscopia se realiza a causa de infertilidad que cuando se hace por ligadura de trompas. Probablemente, en el primer caso se busca con más insistencia, lo que da lugar a un sesgo de información que sobrevalora la asociación. De todos modos, si se corrigiera este sesgo, es muy probable que esta fuerte asociación disminuyera, pero no que desapareciese en su totalidad. Por el contrario, si la asociación que se ha observado es débil, la existencia de un sesgo podría alterar totalmente la interpretación de los resultados. Un estudio positivo puede pasar a ser negativo, y viceversa, según la dirección del sesgo.

Otro elemento que se debe tener en cuenta en la evaluación de la validez interna de un estudio es la existencia de factores de confusión. Se trata de variables que pueden distorsionar la relación entre el factor de estudio y la variable de respuesta. Por tanto, si no se han controlado, o se ha hecho de forma inadecuada, también pueden obtenerse estimaciones distorsionadas del resultado.

**Ejemplo 26.4.** Supongamos un estudio de casos y controles que evalúa si el consumo de alcohol es un factor de riesgo de cáncer de pulmón. Si no se tienen en cuenta otros factores, podría encontrarse una relación positiva entre ambos. Sin embargo, esta conclusión sería errónea, dado que el consumo de alcohol se asocia también al de tabaco, y éste es un factor estrechamente relacionado con el desarrollo de cáncer de pulmón. El consumo de tabaco estaría actuando como un factor de confusión, de forma que, si no se controla su influencia, podría observarse una aparente relación entre el consumo de alcohol y el cáncer de pulmón. En cambio, si se controla en la fase de análisis, esta aparente asociación desaparece e indica que, en realidad, el consumo de alcohol no está asociado con la enfermedad, sino que lo parecía porque estaba relacionado con el tabaco.

## UTILIDAD PRÁCTICA

El hecho de que el resultado de un estudio sea válido no implica necesariamente que sea clínicamente relevante o aplicable a la práctica clínica.

Para poder evaluar la relevancia clínica de un resultado es de gran ayuda calcular su IC, que indica entre qué valores es probable que esté situado el verdadero efecto o asociación, de modo que proporciona una idea de cuál es su verdadera magnitud. Muchos estudios presentan el resultado en forma de una medida relativa (*odds ratio*, riesgo relativo, etc.), pero para evaluar el impacto del resultado en la práctica es preferible utilizar medidas absolutas, como la diferencia de incidencias o la reducción absoluta del riesgo producida por una intervención. Cuando se refiere a intervenciones terapéuticas, una medida absoluta muy útil es el número de personas que es necesario que reciban una determinada intervención para conseguir un éxito o prevenir una muerte o un suceso perjudicial (anexo 5).

Por otro lado, el estudio debe proporcionar información sobre todas las respuestas y los desenlaces relacionados con el factor de estudio. No basta con conocer los beneficios de una intervención sobre una única variable de respuesta, sino que deben

conocerse sus efectos sobre otras variables relevantes, sus efectos secundarios u otros inconvenientes derivados de su administración, etc., de forma que pueda realizarse un balance entre los beneficios y los riesgos derivados del tratamiento.

**Ejemplo 26.5.** Consideremos un estudio que evalúa la eficacia de una intervención farmacológica preventiva que consigue reducir la incidencia de una enfermedad, que es del 2% en el grupo de referencia, hasta el 1%. En términos relativos, se ha reducido el riesgo en un 50% (a la mitad). En términos absolutos, el riesgo se ha reducido en un 1%, lo que significa que es necesario que 100 sujetos reciban la intervención para evitar la aparición de un caso de la enfermedad. Con esta información, puede evaluarse la relación entre los beneficios y los riesgos potenciales de tratar a 100 sujetos.

Supongamos que la misma intervención se aplica a otra población en la que la incidencia es muy

superior, por ejemplo, del 20%. Si la incidencia en el grupo de estudio es del 10%, se ha conseguido la misma reducción relativa del 50%. Sin embargo, en términos absolutos, la diferencia es el 10%, por lo que, por cada 100 pacientes tratados, se evitan 10 casos de enfermedad, o, lo que es lo mismo, es necesario tratar a 10 sujetos para evitar la aparición de un caso.

El ejemplo anterior ilustra cómo las medidas absolutas proporcionan información más útil sobre la relevancia clínica que las relativas, y son más eficaces cuando hay que elaborar recomendaciones prácticas basadas en los resultados de los estudios. Incluso habiendo considerado que un resultado es clínicamente relevante, las recomendaciones que se derivan de él no siempre pueden aplicarse directamente a pacientes o entornos diferentes del estudiado. Por ello, al interpretar los resultados, hay que considerar todos los elementos que pueden limitar su generalización y extrapolación, sobre todo cuando se utilizan para realizar recomendaciones prácticas.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

Strathy JH, Molgaard CA,  
Coulam CB, Melton LJ 3rd.  
Endometriosis and infertility:  
a laparoscopic study of  
endometriosis among fertile  
and infertile women. Fertil  
Steril.. 1982;38:667-72.

## BIBLIOGRAFÍA

Argimon JM. La ausencia  
de significación estadística  
en un ensayo clínico no  
significa equivalencia  
terapéutica. Med Clin (Barc).  
2002;118:701-3.

Argimon JM. Intervalos  
de confianza: algo más que  
una prueba de significación  
estadística Med Clin (Barc).  
2002;118:382-4.

Bland JM. The tyranny of power:  
is there a better way to  
calculate sample size? BMJ.  
2009;339:b3985.

Cobo E, Muñoz P, González JA,  
editors. Estadística para no  
estadísticos: bases para  
interpretar artículos científicos.  
1.ª ed. Barcelona: Elsevier; 2007.

Flather M, Delahunt N, Collinson  
J. Generalizing results of  
randomized trials to clinical  
practice: reliability and cautions.  
Clin Trials. 2006;3(6):508-12.

Glasziou P, Chalmers I, Altman  
DG, Bastian H, Boutron I,  
Brice A, et al. Taking healthcare  
interventions from trial to  
practice. BMJ. 2010;341:c3852.

Higgins JP, Altman DG, Gøtzsche  
PC, Jüni P, Moher D, Oxman  
AD, et al.; Cochrane Bias  
Methods Group; Cochrane  
Statistical Methods Group.  
The Cochrane Collaboration's  
tool for assessing risk of bias  
in randomised trials. BMJ.  
2011;343:d5928.

Keech AC, Pike R, Granger RE,  
Gebski VJ. Interpreting the  
results of a clinical trial. Med J  
Aust. 2007;186(6):318-9.

Paradis C. Bias in surgical research.  
Ann Surg. 2008;248(2):180-8.

## Papel de la estadística

La estadística no es útil exclusivamente para el análisis de los resultados, sino que debe considerarse una parte integrante del método científico que se aplica en diferentes fases de una investigación para facilitar que se alcance el objetivo deseado (*cuadro 27.1*). En la fase de análisis permite evaluar y cuantificar la variabilidad debida al azar. La premisa previa para su uso es que el estudio haya sido diseñado y ejecutado de forma correcta.

Clásicamente, la estadística se diferencia en descriptiva e inferencial. La *estadística descriptiva* permite organizar, presentar y sintetizar la información y es fundamental en la revisión de los datos recogidos en un estudio para asegurar su calidad y la validez del análisis posterior, así como para describir las características de los sujetos estudiados. La *estadística inferencial* permite establecer conclusiones referidas a poblaciones a partir de los resultados obtenidos en muestras. Su aplicación en la fase de análisis tiene dos finalidades principales: evaluar la variabilidad aleatoria y controlar los factores de confusión.

Las técnicas de análisis estadístico pueden utilizarse para explorar conjuntos de datos sin hipótesis previas o bien para confirmar hipótesis de trabajo. Ambas finalidades están vinculadas a la naturaleza de los objetivos del estudio, a la actitud con que el investigador se enfrenta a los datos y a los términos en que deberán interpretarse los resultados. Una hipótesis solamente puede confirmarse mediante un estudio diseñado con el propósito de hacerlo. La exploración implica el rastreo de datos en busca de información, sin objetivos concretos ni hipótesis que hayan gobernado el diseño del estudio. La exploración puede servir para sugerir nuevas hipótesis, pero de ningún modo para contrastarlas, sino que la confirmación deberá obtenerse en un nuevo estudio diseñado para ello. Ambas finalidades pueden

coexistir en un estudio diseñado para confirmar una hipótesis cuando se establecen objetivos secundarios exploratorios. Aunque no están vinculadas a técnicas de análisis concretas, las pruebas de contraste o de significación estadística están dirigidas a confirmar hipótesis, mientras que las técnicas exploratorias son eminentemente gráficas y descriptivas. En este capítulo se abordan, sobre todo, las bases de la utilización de la inferencia estadística para confirmar la hipótesis de trabajo y alcanzar el objetivo del estudio.

### VARIACIONES DEL MUESTREO

En estadística, el término *población* se utiliza para describir todas las posibles observaciones de una determinada variable o todas las unidades sobre las que podría haberse realizado una observación. El significado preciso de este concepto varía en función del contexto en que se utiliza. Puede tratarse de una población de pacientes, de profesionales o de determinaciones de laboratorio, e incluso difiere según la localización geográfica y la fuente de sujetos u observaciones utilizada. A menudo no tiene una realidad física concreta, ya que, por ejemplo, la población de pacientes con infección urinaria no existe como tal, sino que el proceso va apareciendo y desapareciendo en diferentes sujetos en función de múltiples y variados factores.

En sentido amplio, el término *muestra* se refiere a cualquier conjunto específico de sujetos u observaciones procedentes de una población determinada. Para que sea útil y permita aplicar las técnicas estadísticas, se requiere que la muestra tenga un tamaño razonable y sea representativa de la población de la que procede. Un tamaño grande no asegura la representatividad, sino que ésta radica básicamente

### Cuadro 27.1 Fases de una investigación en las que interviene la estadística

- Selección de la variable de respuesta
- Definición de los criterios de selección de la población de estudio
- Elección de la técnica de selección de los sujetos
- Cálculo del número de sujetos necesarios
- Selección de las variables que deben ser medidas
- Medición de las variables (precisión y exactitud)
- Descripción de la muestra de sujetos estudiados
- Estimación de la magnitud del efecto o respuesta observada
- Comparación del efecto observado en diferentes grupos
- Control de los factores de confusión
- Interpretación de los resultados

en que la muestra se haya escogido aleatoriamente y esté libre de sesgos.

Se estudian muestras en lugar de poblaciones por criterios de eficiencia. El propósito fundamental del muestreo es estimar el valor de una determinada variable (parámetro) en la población, a partir de un número menor de observaciones (muestra). Sin embargo, tan sólo se estudia una de las múltiples muestras que podrían seleccionarse de la población de referencia, en cada una de las cuales podría obtenerse un valor diferente, simplemente por azar. Las diferentes técnicas de la estadística inferencial se fundamentan en que esta variabilidad inherente al proceso de muestreo sigue unas leyes conocidas y puede ser cuantificada.

Así, en el caso de una variable cuantitativa, en cada muestra se obtendrá una media diferente. Si se representa gráficamente la distribución de las medias de todas las muestras posibles de un mismo tamaño, se puede comprobar que sigue la ley normal. Esto es cierto si la variable es normal en la población de origen. Pero, aunque no lo sea, la distribución de las medias muestrales tiende a ser normal a medida que su tamaño aumenta, y se acepta que, si las muestras son de más de 30 sujetos, la distribución de sus medias es normal.

Esta distribución de medias muestrales tiene dos características que la hacen especialmente interesante. La primera es que su media es la media de la población de la que proceden las muestras; es decir, las medias muestrales se distribuyen normalmente alrededor de la media poblacional desconocida que se quiere estimar. La segunda es que la desviación estándar de dicha distribución tiene el valor  $\sigma/\sqrt{n}$ , donde  $\sigma$  es la desviación estándar de la población de origen y  $n$  el tamaño de la muestra. Este valor es conocido como *error estándar*

de la media (EEM), y mide la dispersión de las medias muestrales respecto de la media poblacional. No debe confundirse con la desviación estándar, cuyos valores miden la dispersión de los valores de la variable (no de las medias muestrales) en los sujetos de la población ( $\sigma$ ) o de la muestra (DE<sub>s</sub>), respectivamente (fig. 27.1). Puede deducirse fácilmente que el EEM disminuye cuando aumenta el tamaño de la muestra  $n$ , lo que explica el hecho de que las muestras grandes estimen el valor poblacional con mayor precisión.

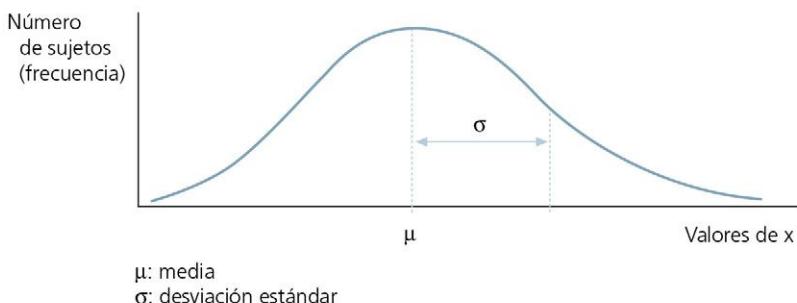
En el caso de las variables cualitativas, la distribución de la proporción sigue la ley binomial. Sin embargo, cuando los productos  $n \cdot p$  y  $n \cdot (1 - p)$  son superiores a 5, la distribución se asemeja bastante a la normal. Así, las proporciones observadas en las muestras se distribuyen alrededor de la verdadera proporción poblacional, y la dispersión de esta distribución se mide mediante el *error estándar de la proporción* (EEP), cuyo valor es  $\sqrt{p(1-p)/n}$ , y tiene características similares al EEM.

## ESTIMACIÓN DE UN PARÁMETRO POBLACIONAL: INTERVALO DE CONFIANZA

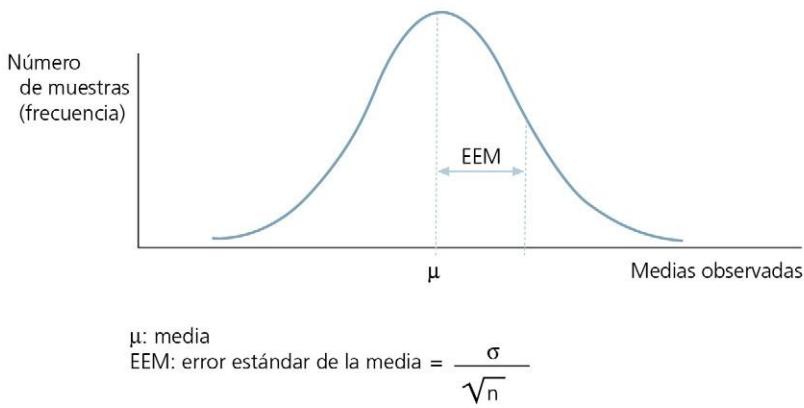
Un objetivo frecuente en la investigación médica es estimar un parámetro poblacional a partir de los valores que la variable de interés adopta en los individuos de una muestra. Si la variable es cuantitativa, la media (m) y la desviación estándar (DE) observadas en la muestra son la mejor estimación disponible de los verdaderos valores de los parámetros poblacionales. Pero ¿cuáles serían los resultados si se repitiera el estudio en múltiples ocasiones?

**Ejemplo 27.1.** Supongamos que en una muestra de 60 sujetos se observa una media de presión arterial sistólica (PAS) de 150 mmHg, con una DE de 20 mmHg, y que se desea conocer el verdadero valor de la PAS media en la población de referencia. En principio, el valor más probable es la *estimación puntual* obtenida en la muestra (150 mmHg), pero, dado que si se hubiera estudiado una muestra diferente probablemente se habría obtenido un resultado distinto, se necesita una medida de la precisión de esta estimación, lo que se hace mediante el cálculo del llamado *intervalo de confianza* (IC) (cuadro 27.2). Habitualmente se trabaja con una confianza del 95%, es decir, con un valor  $\alpha$  del 5%, que corresponde a un valor  $Z_{\alpha}$  de 1,96. Aplicando la fórmula, se obtendría un IC del 95%, que sería aproximadamente de  $150 \pm 5$  mmHg, lo que significa que hay un 95% de confianza de que el valor medio de la PAS de la población de referencia se encuentre entre 145 y 155 mmHg.

Distribución de los valores de una variable  $x$  medidos en los individuos de una población



Distribución de las medias observadas en las múltiples muestras de tamaño  $n$  que pueden obtenerse de una población de media  $\mu$  y desviación estándar  $\sigma$



**Figura 27.1** Diferencia entre desviación estándar y error estándar de la media.

De forma similar se calcularía el IC en el caso de una variable cualitativa (ver cuadro 27.2).

El IC proporciona mucha más información que la estimación puntual, ya que permite evaluar la precisión con que se ha estimado el parámetro poblacional; es decir, entre qué límites se tiene una determinada confianza de que esté situado su verdadero, pero desconocido, valor. Si se repitiera el estudio en 100 ocasiones, el IC incluiría el verdadero valor en 95 de ellas. Sin embargo, no puede descartarse totalmente que el estudio corresponda a una de las cinco ocasiones restantes.

De las fórmulas del cuadro 27.2 se deduce que un aumento del número de sujetos conduce a un IC más estrecho y a un aumento de la precisión de la estimación. La amplitud del IC depende también del grado de confianza que se utilice, y aumenta a medida que se incrementa la confianza deseada.

En el cálculo del IC se asume que se ha estudiado una muestra aleatoria y representativa de la población

de referencia. Al interpretarlo, hay que tener en cuenta la posibilidad de la existencia de otras fuentes de error no debidas al azar (errores sistemáticos o sesgos). Si éstas existen, o si la muestra no es aleatoria, el error de la estimación puede ser mayor que el sugerido por la amplitud del intervalo.

Siempre que se realizan inferencias sobre parámetros poblacionales a partir de criterios estadísticos muestrales, los resultados deben expresarse como IC, y no sólo como estimaciones puntuales, para poder valorar la precisión de la estimación.

## CONTRASTE DE HIPÓTESIS

Aunque la situación es similar a la anterior, dado que se pretende descubrir algo sobre las poblaciones a partir del estudio de muestras, las pruebas de contraste de hipótesis, o de significación estadística,

### Cuadro 27.2 Cálculo del intervalo de confianza (IC) en la estimación de un parámetro poblacional

IC de una media (variable cuantitativa)\*:

$$\bar{x} \pm (Z\alpha \cdot EEM)$$

$$\text{siendo } EEM = \frac{s}{\sqrt{n}}$$

IC de una proporción (variable cualitativa)\*\*:

$$p \pm (Z\alpha \cdot EEP)$$

$$\text{siendo } EEP = \sqrt{\frac{p \cdot (1-p)}{n}}$$

$\bar{x}$ : media observada en la muestra.

s: desviación estándar observada en la muestra.

n: número de individuos de la muestra.

EEM: error estándar de la media.

p: proporción observada en la muestra.

EEP: error estándar de la proporción.

$Z\alpha$ : valor de la variable normal tipificada correspondiente al valor  $\alpha$  para un valor de confianza  $(1 - \alpha)$ .

\* Este cálculo se basa en la distribución normal. El valor de  $Z\alpha$  para un IC del 95% es 1,96. Para muestras de tamaño inferior a 30 individuos, este valor debe sustituirse por el de la distribución de la t de Student-Fisher para  $(n - 1)$  grados de libertad.

\*\* Las variables cualitativas no presentan una distribución normal. Las fórmulas de la tabla se basan en una aproximación a la normalidad, aplicable cuando los productos  $np$  y  $n(1 - p)$  son mayores de 5. En caso contrario, debe aplicarse una corrección al valor de  $Z\alpha$ .

valoran la variabilidad debida al azar de forma diferente a la estimación de los parámetros.

Existen diversas pruebas estadísticas aplicables en diferentes situaciones en función del número de grupos que se comparan, la escala de medida de las variables, el número de sujetos analizados, etc. (anexo 9). En este capítulo se presentan los fundamentos comunes a todas ellas, haciendo especial énfasis en los aspectos relacionados con la interpretación de los resultados obtenidos.

Supongamos que existe interés en comparar dos tratamientos (un diurético D y el tratamiento estándar E) para determinar cuál de ellos es el más eficaz en el control de la presión arterial. Se diseña un ensayo clínico, distribuyendo aleatoriamente 60 pacientes hipertensos en dos grupos, cada uno de los cuales recibe uno de los tratamientos. A los 3 meses, el porcentaje de individuos controlados en cada grupo es del 70 y el 50%, respectivamente. ¿Qué conclusión puede obtenerse a la vista de estos

resultados? ¿Hasta qué punto es posible que se deban simplemente al azar (variaciones del muestreo) y que en realidad no exista ninguna diferencia de eficacia entre ambas intervenciones?

### Hipótesis nula e hipótesis alternativa

Siguiendo con el ejemplo anterior, el análisis parte de la hipótesis de que no existen diferencias entre los porcentajes de hipertensos controlados observados en ambos grupos. La prueba de significación estadística intentará rechazar esta hipótesis, conocida como *hipótesis nula* ( $H_0$ ). Si se consigue, se aceptará la *hipótesis alternativa* ( $H_a$ ), según la cual existen diferencias entre ambos grupos.

El primer paso es formular la  $H_0$ . A continuación se calcula, mediante la prueba estadística más adecuada, la probabilidad de que los resultados observados puedan deberse al azar, en el supuesto de que la  $H_0$  sea cierta. En otras palabras, la probabilidad de que, a partir de una población de referencia, puedan obtenerse dos muestras que presenten unos valores tan diferentes como los observados simplemente por azar. Esta probabilidad es el *grado de significación estadística*, que suele representarse con la letra  $p$ . En tercer lugar, basándose en esta probabilidad, se decide si se rechaza o no la  $H_0$ . Cuanto menor sea la  $p$ , es decir, cuanto menor sea la probabilidad de que el azar pueda haber producido los resultados observados, mayor será la evidencia en contra de la  $H_0$  y, por tanto, mayor será la tendencia a concluir que la diferencia existe en la realidad.

Supongamos que, en el ejemplo, una vez aplicada la prueba estadística adecuada, se obtiene un valor de  $p$  aproximadamente de 0,10. Esto significa que, si la  $H_0$  fuera cierta, la probabilidad de que el azar pueda producir unos resultados como los observados es del 10%, es decir, que existe un 10% de probabilidad de que dos muestras de 30 sujetos obtenidas de una misma población presenten unos porcentajes del 70 y el 50% sólo por variabilidad aleatoria.

Para decidir si se rechaza o no la  $H_0$  debe fijarse previamente un valor de  $p$  por debajo del cual se considera que se dispone de la suficiente evidencia en su contra para rechazarla. Este valor se conoce como *valor de significación estadística*  $\alpha$ . De forma arbitraria, y por convenio, se fija habitualmente en el 5% (0,05).

Dado que el valor de  $p$  obtenido en el ejemplo es de 0,10, superior al valor de significación de 0,05, se considera que la probabilidad de haber obtenido estos resultados por azar es demasiado elevada y que, por tanto, no se dispone de la suficiente evidencia para rechazar la  $H_0$ . De este modo, se concluye que la diferencia observada en el porcentaje de pacientes controlados no es estadísticamente significativa. No significa que no exista diferencia en los porcentajes

de ambos grupos, sino que no se ha encontrado la suficiente evidencia para decir que son diferentes.

Supongamos ahora que en el grupo que recibió el tratamiento E sólo se hubieran controlado 12 pacientes a los 3 meses, lo que supondría un porcentaje observado en este grupo del 40%. Si se repiten los cálculos, se obtiene un valor de  $p < 0,02$ . Como este valor es inferior a 0,05, se considera que la diferencia observada es estadísticamente significativa, ya que es poco probable ( $p < 5\%$ ) que el azar pueda haber producido estos resultados. La respuesta a la pregunta de si esta diferencia se debe al nuevo tratamiento D dependerá del diseño y la ejecución correctos del estudio. El valor de  $p$  sólo informa de la existencia de una diferencia entre ambos grupos, y de que muy probablemente no se deba al azar, pero no informa sobre la causa de dicha diferencia.

El valor de  $p$  no es una medida de la fuerza de la asociación. Un estudio en el que se obtenga un valor de  $p < 0,001$  no quiere decir que la asociación encontrada sea más fuerte (o la diferencia más importante) que la de otro estudio en que sea de 0,04. Sólo quiere decir que es más improbable que su resultado sea por azar.

No hay que ser excesivamente rígido en el límite del valor de significación. Un valor de  $p = 0,048$  es estadísticamente significativo con el umbral del 5%, y uno de 0,052, en cambio, no lo es, pero en ambos casos la probabilidad de observar el resultado por azar es prácticamente la misma, y muy próxima al 5%. Por ello, es conveniente indicar el valor de  $p$  al dar los resultados, sobre todo si es próximo al valor de significación, en lugar de limitarse a decir si existe o no significación estadística. De esta forma, el lector podrá valorar adecuadamente los resultados.

## Pruebas unilaterales y pruebas bilaterales

En ocasiones, lo que interesa no es determinar si existen diferencias entre dos tratamientos, sino evaluar si un nuevo fármaco es mejor que otro. En este caso, la hipótesis alternativa no es que D y E difieran, sino que D es mejor que E. Por tanto, la  $H_0$  que se va a contrastar es que D no difiere o es peor que E. Dado que sólo interesa un sentido de la comparación, se habla de pruebas unilaterales, o de una cola.

¿Cómo afecta este hecho a la prueba de significación? No es la prueba en sí misma la que se ve afectada. El cálculo es idéntico al anterior. Lo que se modifica es el valor de  $p$ . Como la distribución de Z sigue la ley normal y, por tanto, es simétrica, en las pruebas unilaterales el valor de  $p$  corresponde a la mitad del valor  $\alpha$ , dado que sólo se está interesado en uno de los extremos.

## Error $\alpha$ y error $\beta$

En estadística no puede hablarse de certeza absoluta. Sea cual sea la decisión que se tome respecto a la hipótesis nula, se corre un cierto riesgo de equivocarse (fig. 27.2).

La realidad no es conocida, ya que, si lo fuera, no sería necesario realizar el estudio. Si no se rechaza la  $H_0$ , y ésta es cierta, no se comete ningún error. Si se rechaza y es falsa, tampoco. Pero ¿qué pasa en las otras dos situaciones?

En un estudio puede concluirse que existe una diferencia cuando en realidad no la hay. Es decir, puede rechazarse la  $H_0$  cuando en realidad es cierta. Si esto ocurre, la decisión es incorrecta y se comete un error, conocido como error tipo I o error  $\alpha$ . La probabilidad de cometer este error es la de que, si se concluye que existe una diferencia significativa, ésta sea en realidad debida al azar. Si se hace un símil entre una prueba estadística y una diagnóstica, equivale a la probabilidad de obtener un resultado falso positivo. Esto es precisamente lo que mide el valor de  $p$ , o grado de significación estadística de la prueba.

Si, por el contrario, se concluye que la diferencia no es estadísticamente significativa, es decir, si no se rechaza la  $H_0$ , puede ocurrir que la hipótesis sea falsa y que, en realidad, exista una diferencia entre ambos grupos, en cuyo caso se cometerá otro tipo de error, llamado error tipo II o  $\beta$ . Utilizando el símil con la prueba diagnóstica, equivale a la probabilidad de obtener un resultado falso negativo. Su valor complementario  $1 - \beta$ , denominado *potencia*, o *poder estadístico*, indica la capacidad que tiene la prueba para detectar una diferencia que existe en la realidad. Lógicamente, cuanto mayor es la diferencia y más elevado el número de individuos estudiados, mayor capacidad existe para detectarla; es decir, el poder estadístico es mayor y, por tanto, la probabilidad de cometer un error  $\beta$  es menor.

Existe una interdependencia entre el grado de significación  $p$ , la potencia estadística, el número de individuos estudiados y la magnitud de la diferencia observada. Conociendo tres de estos parámetros, puede calcularse el cuarto. Así, antes de iniciar un estudio, puede calcularse el número de sujetos necesario, fijando a priori el grado de significación, la potencia estadística y el valor de la diferencia que quiere detectarse. De igual modo, si una vez acabado el estudio se concluye que no se ha encontrado una diferencia estadísticamente significativa, dado que  $n$ ,  $p$  y la diferencia observada son conocidas, puede calcularse el poder estadístico. No es lo mismo concluir que no se ha encontrado una diferencia estadísticamente significativa entre dos tratamientos cuando se tiene una probabilidad del 90% de haberla

		Realidad	
		No existe diferencia ( $H_0$ cierta)	Existe diferencia ( $H_0$ falsa)
Resultado de la prueba estadística	Diferencia significativa (rechazo de $H_0$ )	Error tipo I ( $\alpha$ )	No error
	Diferencia no significativa (no rechazo de $H_0$ )	No error	Error tipo II ( $\beta$ )

$H_0$ : hipótesis nula;  $H_a$ : hipótesis alternativa; A: distribución de la diferencia cuando la  $H_0$  es cierta (media = O); B: distribución de la diferencia cuando la  $H_a$  es cierta (media desconocida = D); R: resultado (diferencia) observado en el estudio.

**Figura 27.2** Resultados de una prueba de significación estadística.

detectado si hubiera existido ( $\beta = 0,10$ ), que cuando esta probabilidad es sólo del 20% ( $\beta = 0,20$ ).

### ¿Diferencia estadísticamente significativa o clínicamente relevante?

Un resultado estadísticamente significativo no implica que sea clínicamente relevante. El valor de  $p$  no mide la fuerza de la asociación. Pueden obtenerse valores pequeños de  $p$  (resultados estadísticamente significativos) simplemente estudiando un número elevado de sujetos. Al aumentar el tamaño de la muestra se incrementa el poder estadístico para detectar incluso pequeñas diferencias.

**Ejemplo 27.2.** Supongamos un estudio en que se comparan dos fármacos, D y E, para el tratamiento de la hipertensión arterial. El porcentaje de pacientes controlados en el grupo que ha recibido el fármaco D es del 70%, y en el que ha sido tratado con E, del 65%. En la tabla 27.1 se aprecia que la conclusión de si esta diferencia es o no estadísticamente significativa depende del número de individuos. Estudiando 30 pacientes en cada grupo, se concluiría que la dife-

rencia no es estadísticamente significativa. A medida que aumenta el tamaño de los grupos, aumenta el valor de  $Z$  y disminuye el de  $p$ . Al estudiar 700 pacientes en cada grupo ya se alcanza el nivel de significación estadística 9 del 0,05. Y si se estudiaran 2.000 individuos, el valor de  $p$  sería menor de 0,001.

La diferencia que se considera clínicamente relevante depende de su magnitud y de otros factores, como la frecuencia y la gravedad de los efectos secundarios de ambos fármacos, la facilidad de administración o su coste económico.

### Estimación frente a significación estadística

Al analizar los resultados de un estudio, los investigadores están interesados no sólo en saber si una diferencia o una asociación son estadísticamente significativas, sino también en determinar su magnitud. La diferencia observada en el estudio es la mejor estimación puntual de dicha magnitud. Pero, dado que si se repitiera el estudio con otras muestras podrían observarse resultados de diferente magnitud, hay que

**Tabla 27.1** Influencia del número de sujetos estudiados sobre el grado de significación estadística de la comparación de los porcentajes de pacientes controlados con dos tratamientos: D (70%) y E (65%) (ejemplo 27.2)

Número de sujetos estudiados por grupo	Valor de Z*	Valor de p
30	0,41	0,68
100	0,76	0,45
200	1,06	0,29
500	1,69	0,09
700	2,00	0,05
1.000	2,39	< 0,02
2.000	3,38	< 0,001

\*Valor de la Z de comparación de dos proporciones.

calcular un intervalo del que se tenga una determinada confianza de que contiene la verdadera magnitud del parámetro de interés, tanto si se trata de una diferencia (cuadros 27.3 y 27.4) como de una medida de asociación (anexo 2). Existen fórmulas para el cálculo de los IC en cualquier situación.

Cuando se utiliza como medida del efecto una diferencia, si el IC del 95% incluye el valor 0, que es el que corresponde a la  $H_0$  (ausencia de diferencia entre ambos grupos), se concluirá que el resultado no es estadísticamente significativo, ya que no puede descartarse que ése sea el verdadero valor. Si, por el contrario, el IC excluye este valor 0, se concluirá que la diferencia observada es estadísticamente significativa, ya que puede descartarse dicho valor (con un riesgo de error inferior al 5%). Por tanto, el IC, además de indicar si la diferencia es o no estadísticamente significativa, permite conocer entre qué límites es probable que se encuentre la verdadera diferencia, lo que es muy útil en la interpretación de los resultados.

**Ejemplo 27.3.** Supongamos un estudio que compara la eficacia de dos tratamientos A y B en dos grupos de 30 pacientes. Se observa una diferencia en el porcentaje de éxitos del 20% (70 – 50%) a favor del tratamiento B, que no es estadísticamente significativa ( $p=0,12$ ). El IC del 95% de la diferencia entre los dos tratamientos es:

$$\text{IC } 95\%: 0,2 \pm 0,24; \text{ es decir, de } -4 \text{ a } 44\%$$

Con un 95% de confianza, la verdadera magnitud de la diferencia está en el intervalo comprendido entre un 4% a favor del tratamiento A y un 44% a favor de B. Dado que una diferencia del 0% es posible, no puede descartarse que éste sea su verdadero valor,

por lo que el resultado no es estadísticamente significativo. En cambio, el IC informa, además, que también son posibles grandes diferencias a favor de B, y que son improbables grandes diferencias a favor de A. Aunque los resultados siguen sin ser concluyentes, se dispone de más información para interpretarlos de forma adecuada. El IC cuantifica el resultado encontrado y provee un rango donde es muy probable que se encuentre el valor real que se está buscando.

**Ejemplo 27.4.** Supongamos un estudio que compara la eficacia de dos tratamientos A y B en dos grupos de 80 pacientes. Se observa una diferencia del 5% (65 – 60%) a favor del tratamiento B, que es estadísticamente significativa ( $p=0,04$ ). El IC del 95% de esta diferencia es:

$$\text{IC } 95\%: 0,05 \pm 0,047; \text{ es decir, de } 0,3 \text{ a } 9,7\%$$

Al excluir el valor 0% se concluye que la diferencia es estadísticamente significativa. El IC informa que la diferencia es de pequeña magnitud, como máximo de aproximadamente un 10% a favor de B.

Los IC tienen otra ventaja adicional, y es que expresan los resultados en las unidades en que se han realizado las mediciones, lo que permite al lector considerar críticamente su relevancia clínica.

Al diseñar un estudio, los autores establecen la magnitud mínima de la diferencia que consideran de relevancia clínica, en función de la cual han calculado el tamaño necesario de la muestra. Al acabar el estudio, la interpretación del resultado observado y de su IC debe tener en cuenta también esta magnitud. Además de determinar si el IC del 95% excluye el valor 0 para saber si el resultado es estadísticamente significativo, también debe determinarse

### Cuadro 27.3 Cálculo del intervalo de confianza (IC) de la diferencia entre dos proporciones

IC de la diferencia de dos proporciones\*

a) Muestras independientes

$$(P_A - P_B) \pm Z\alpha \cdot EED$$

$$\text{siendo EED} = \sqrt{\frac{P_A \cdot (1-P_A)}{n_A} + \frac{P_B \cdot (1-P_B)}{n_B}}$$

b) Muestras apareadas

$$(P_A - P_B) \pm Z\alpha \cdot EED$$

$$\text{siendo EED} = \frac{1}{n} \sqrt{b + c - \frac{(b - c)^2}{n}}$$

$P_A, P_B$ : proporciones observadas en las muestras A y B.

$n_A, n_B$ : número de sujetos de las muestras A y B.

b, c: número de casos que presentan valores diferentes en ambas mediciones (series apareadas).

n: número total de casos.

EED: error estándar de la diferencia.

$Z\alpha$ : valor de la variable normal tipificada correspondiente al valor  $\alpha$ , para un nivel de confianza  $(1 - \alpha)$ .

\* Las variables cualitativas no presentan una distribución normal. Las fórmulas de la tabla corresponden a una aproximación a la normalidad, aplicable cuando todos los productos  $n \cdot P_A \cdot n \cdot (1 - P_A)$ ,  $n \cdot P_B$  y  $n \cdot (1 - P_B)$  son mayores de 5. En caso contrario, deben aplicarse correcciones en el valor de  $Z\alpha$ .

si incluye o excluye el valor de la mínima diferencia de relevancia clínica, para poder evaluar si el estudio es concluyente acerca de la existencia de una diferencia clínicamente importante.

**Ejemplo 27.5.** En la figura 27.3 se presentan seis posibles resultados de un estudio que compara dos grupos y utiliza como medida del resultado la diferencia entre los porcentajes observados en cada uno de ellos. Supongamos que los investigadores establecieron a priori que la mínima diferencia de relevancia clínica era del 20%.

**Situación A.** Se observa una diferencia del 10% (IC 95%: -5 a +25%), que no es estadísticamente significativa, ya que el IC incluye el valor 0. Pero el valor 20% también es un valor posible, ya que está situado en el interior del IC. Se trata, por tanto, de un resultado que no permite descartar ninguna conclusión.

**Situación B.** En esta ocasión se observó una diferencia entre los grupos del 5% (IC 95%: -5 a +15%). No es estadísticamente significativa, ya que el IC incluye

### Cuadro 27.4 Cálculo del intervalo de confianza (IC) de la diferencia entre dos medias

IC de la diferencia de dos medias\*

a) Muestras independientes

$$(\bar{x}_A - \bar{x}_B) \pm Z\alpha \cdot EED$$

$$\text{siendo EED} = s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

$$y s = \sqrt{\frac{s_A^2(n_A - 1) + s_B^2(n_B - 1)}{n_A + n_B - 2}}$$

b) Muestras apareadas

$$\bar{x}_d \pm Z\alpha \cdot EEx_d$$

$\bar{x}_A, \bar{x}_B$ : medias observadas en las muestras A y B.

$s_A, s_B$ : desviaciones estándar observadas en las muestras A y B.

$n_A, n_B$ : número de sujetos de las muestras A y B.

EED: error estándar de la diferencia.

$\bar{x}_d$ : media de las diferencias de las dos mediciones en cada individuo (series apareadas).

$EEx_d$ : error estándar de la media de las diferencias individuales.

$Z\alpha$ : valor de la variable normal tipificada correspondiente al valor  $\alpha$ , para un nivel de confianza  $(1 - \alpha)$ .

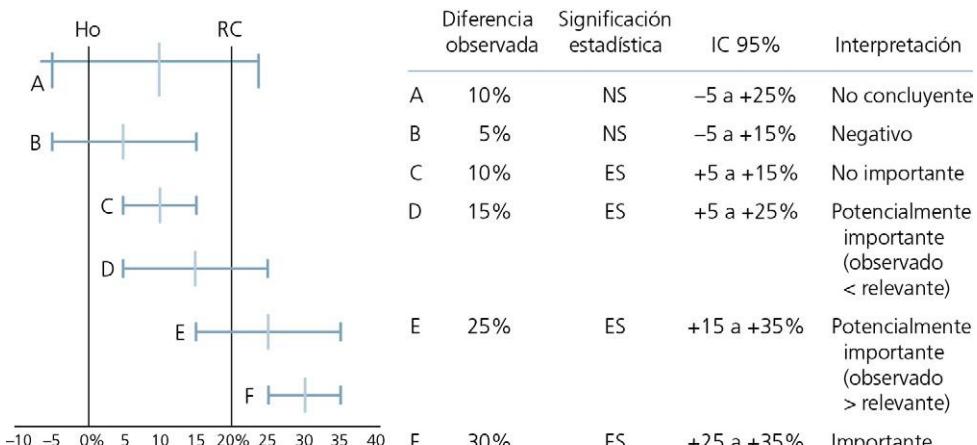
\* El cálculo se basa en la distribución normal. El valor de  $Z\alpha$  para un IC del 95% es 1,96. Para muestras de tamaño inferior a 30 individuos, este valor debe sustituirse por el de la t de Student para  $(n - 1)$  grados de libertad. Así mismo, el cálculo requiere que no existan diferencias significativas entre las desviaciones estándar de ambas muestras.

el valor 0, pero excluye el valor 20%. Así pues, el resultado es negativo en el sentido de que no puede descartarse que los grupos sean iguales, pero, aunque fueran diferentes, es muy improbable que la diferencia sea mayor del 15%. Por tanto, puede descartarse que exista una diferencia de relevancia clínica.

**Situación C.** La diferencia observada del 10% (IC 95%: +5 a +15%) es estadísticamente significativa, ya que el IC excluye el valor 0. Dado que el límite superior del IC no alcanza el valor 20%, puede concluirse que existe una diferencia, pero que ésta no es de relevancia clínica.

**Situación D.** La diferencia observada es del 15% (IC 95%: +5 a +25%), estadísticamente significativa y potencialmente importante, ya que el valor 20% es un valor posible. Así pues, el estudio no es del todo concluyente.

**Situación E.** Similar a la anterior. Aunque la diferencia observada es mayor del 20%, el resultado no es del todo concluyente, ya que el límite inferior del



**Figura 27.3** Posibles resultados de un estudio que compara dos grupos y utiliza como medida del efecto la diferencia entre los porcentajes observados. Se considera que la mínima diferencia de relevancia clínica es del 20% (ejemplo 27.5).

IC está por debajo del 20% e indica que la verdadera diferencia podría ser inferior a este valor.

*Situación F.* La diferencia observada es del 30%. Dado que el límite inferior del IC es mayor del 20%, este resultado es estadísticamente significativo y concluyente a favor de la existencia de una diferencia de relevancia clínica.

Cuando se utiliza una medida relativa del efecto (riesgo relativo, *odds ratio*, etc.), la  $H_0$  de igualdad entre los grupos corresponde al valor 1, ya que se trata de un cociente y no de una diferencia. Por tanto, cuando el IC del 95% de una medida relativa incluye el valor 1, el resultado no es estadísticamente significativo.

Aunque las pruebas de significación continúan siendo los procedimientos estadísticos más empleados, las ventajas de la utilización complementaria de los IC en el análisis e interpretación de los resultados, tanto si el objetivo es la estimación de parámetros como el contraste de una hipótesis, hacen que cada vez se les conceda una mayor importancia y las directrices para la publicación de estudios científicos recomiendan que se presenten los resultados principales con su correspondiente IC en lugar de informar solamente del valor de  $p$ .

si los grupos difieren por alguna variable, cuando se comparan diferentes variables de respuesta o cuando se analizan diversos subgrupos de sujetos. Efectuar comparaciones múltiples tiene dos grandes inconvenientes:

- La realización de pruebas para cada variable por separado ignora el hecho de que muchas de ellas pueden estar relacionadas entre sí, de forma que el resultado de una prueba estadística determinada puede estar influido por diferencias en la distribución de otras variables relacionadas.
- Si cada una de las pruebas estadísticas se realiza con el nivel de significación prefijado del 5%, en promedio, 5 de cada 100 comparaciones pueden resultar significativas sólo por azar. Al realizar múltiples comparaciones, aumenta la probabilidad de obtener algún resultado estadísticamente significativo que no refleje una diferencia real.

Para estimar la probabilidad de obtener un resultado significativo por azar tras realizar un número  $n$  de pruebas estadísticas, cada una de ellas con el nivel de significación  $\alpha$ , puede usarse la llamada *desigualdad de Bonferroni*, de la que se deriva la siguiente fórmula:

$$Pr = 1 - (1 - \alpha)^n$$

donde  $Pr$  representa la probabilidad de encontrar un resultado significativo.

## Comparaciones múltiples

En cualquier estudio se realizan habitualmente múltiples comparaciones; por ejemplo, cuando se evalúa

**Ejemplo 27.6.** Supongamos que se realizan 8 comparaciones independientes, cada una de ellas con el nivel de significación de 0,05. La probabilidad de que alguna de estas comparaciones conduzca a un resultado significativo simplemente por azar es:

$$\text{Pr} = 1 - (1 - 0,05)^8 = 0,336$$

es decir, existe un 33% de probabilidades de cometer algún error  $\alpha$ , y no el 5% con el que se deseaba trabajar.

Puede utilizarse como aproximación el producto del valor de significación por el número de pruebas realizadas ( $n \cdot \alpha$ ). En el ejemplo, este valor sería 0,40, que, como puede observarse, sobreestima ligeramente dicha probabilidad.

La solución más adecuada a este problema es reducir en lo posible el número de comparaciones que realizar, y aplicarlas solamente para contrastar hipótesis previas, y asignar prioridades antes del análisis, para decidir cuál es la comparación principal y cuáles las secundarias, de forma que estas últimas se utilicen para matizar la respuesta a la primera.

Otra solución es dividir el valor de significación deseado por el número de pruebas que se van a realizar y obtener así un valor de significación corregido para aplicarlo a cada una de las pruebas. Este procedimiento se conoce como *corrección de Bonferroni*. En el ejemplo 27.6, si se desea mantener el valor de significación global del 0,05, cada una de las pruebas deberá realizarse con el valor corregido de  $0,05/8 = 0,00625$ .

La corrección de Bonferroni asume que las comparaciones son independientes, por lo que no resulta adecuada cuando los análisis están mutuamente asociados.

**Ejemplo 27.7.** Supongamos que un estudio utiliza como variable principal la variación en los valores de hemoglobina, y se observa una reducción estadísticamente significativa. Al analizar también los valores de hematocrito, también se obtiene una disminución significativa, pero al aplicar la corrección de Bonferroni, la reducción de la hemoglobina deja de ser estadísticamente significativa. Este hecho no parece lógico, ya que realmente la variación del hematocrito estaría confirmando el resultado principal. El problema es que las comparaciones no son independientes, ya que se trata de parámetros muy correlacionados (como el hematocrito y la hemoglobina), de manera que si el resultado es significativo en uno de ellos, es probable que lo sea también en el otro.

El inconveniente principal de la corrección de Bonferroni es que protege excesivamente contra la posibilidad de rechazar de forma errónea alguna de las hipótesis nulas a costa de disminuir la poten-

cia de la prueba. Dado que el intento de ajustar el efecto de la multiplicidad de análisis podría requerir establecer niveles de significación estadística absurdamente pequeños, parece más razonable adoptar una actitud conservadora al interpretar los resultados de análisis múltiples y valorar su consistencia entre las diferentes variables. Una vez más, lo ideal sería definir a priori la variable principal y unas pocas variables secundarias como protección frente a este problema. No sería lógico que la interpretación de los resultados observados fuera diferente según el número de pruebas estadísticas que se llevan a cabo. Si, por ejemplo, se comparan dos pautas de quimioterapia y se observa una diferencia en las tasas de remisión, el hecho de considerarla o no estadísticamente significativa no debería depender de si se han evaluado además otros parámetros, como la supervivencia, las complicaciones o la calidad de vida.

De hecho, existen opiniones contrarias al uso rutinario de este ajuste, fundamentalmente cuando se trata de múltiples resultados, aunque parece más adecuado su uso cuando se comparan diferentes subgrupos o se realizan análisis secuenciales. Una alternativa más adecuada es el empleo de técnicas multivariantes que tomen en consideración las relaciones que existen entre las variables que se comparan.

Un problema similar puede presentarse en estudios con un seguimiento prolongado, cuando el investigador decide analizar sus datos repetidamente a medida que el estudio progresá. El uso del valor de significación estadística según el método habitual no resulta apropiado, ya que el cálculo convencional asume que el tamaño de la muestra es un valor fijo, y que los resultados del estudio se analizarán una única vez con los datos de todos los sujetos. Por tanto, si un investigador analiza los datos en diversas ocasiones, la probabilidad de alcanzar un resultado significativo es mayor que la deseada. Cuando se efectúa este tipo de estudios, lo mejor es utilizar un diseño secuencial, en el que no es necesario asumir que el análisis se realizará una sola vez, y en el que las normas para finalizar el estudio permiten la evaluación continuada de los datos.

## Análisis multivariante

Hay veces en las que interesa considerar la influencia de más de dos variables simultáneamente. Ello requiere técnicas sofisticadas, basadas en modelos matemáticos complejos, agrupadas bajo el nombre genérico de *análisis multivariante*.

Existen múltiples técnicas estadísticas multivariantes. En investigación clínica y epidemiológica, las más utilizadas son las que analizan la relación entre una *variable dependiente* (variable de respuesta) y un grupo

de *variables independientes* (factor de estudio y variables que controlar). Estas técnicas implican la construcción de un modelo matemático. La elección de un modelo u otro dependerá del diseño del estudio, la naturaleza de las variables y las interrelaciones entre el factor de estudio, la variable de respuesta y las restantes variables incluidas en el modelo (anexo 9).

Estas técnicas pueden aplicarse con las siguientes finalidades:

- Proporcionar una estimación del efecto de una variable independiente principal (factor de estudio) sobre una dependiente (variable de respuesta), *ajustada* por un conjunto de factores o variables independientes (potenciales factores de confusión y modificadores del efecto).

**Ejemplo 27.8.** Supongamos un ensayo clínico aleatorio que evalúa el efecto de un nuevo fármaco (variable independiente principal) sobre la colesterolemia (variable dependiente), controlando la influencia de la edad y el sexo de los sujetos (variables independientes).

- Describir la relación que existe entre un conjunto de variables independientes (sin identificar una de ellas como principal) y una variable dependiente, y la contribución de cada una de ellas a la relación.

**Ejemplo 27.9.** Supongamos un estudio transversal que pretende evaluar, de entre un conjunto de variables independientes (colesterol, edad, sexo, glucemia, etc.), cuáles están asociadas con una variable dependiente (cifras de presión arterial).

- Predecir el valor de una variable dependiente en función de los valores que toma un conjunto de variables independientes.

**Ejemplo 27.10.** Supongamos un estudio de cohortes en el que se desea obtener una ecuación que prediga el riesgo cardiovascular (variable dependiente) en función de la edad, el sexo, las cifras de presión arterial, el consumo de tabaco y las cifras de colesterol de los sujetos (variables independientes).

Las técnicas que relacionan un conjunto de variables independientes con una dependiente derivan del modelo de regresión lineal, y pueden clasificarse a partir de la escala de medida de la variable dependiente (anexo 9).

Cuando la finalidad del análisis es predictiva o descriptiva de las relaciones entre variables, lo que interesa es obtener una ecuación o un modelo lo más sencillo posible, de forma que la selección de las variables que formarán el modelo se basa en criterios

de significación estadística. En cambio, cuando la finalidad es estimar el efecto del factor de estudio controlando determinados factores de confusión, no puede confiarse en que la selección automática incluya las variables que interesa controlar, ya que no siempre están asociadas de forma estadísticamente significativa, por lo que se utilizan otros procedimientos que obligan a la selección por parte del investigador de las variables que desea controlar.

La correcta utilización de la estadística en el análisis de los datos es fundamental para poder interpretarlos de forma adecuada y obtener conclusiones válidas. Por ello, una vez realizado el análisis multivariante, hay que evaluar la adecuación del modelo obtenido (bondad de ajuste), ya que, por ejemplo, una ecuación puede ser estadísticamente significativa, pero predecir con poca fiabilidad el riesgo de enfermar de un sujeto. Debe prestarse atención a las medidas globales de bondad del modelo, así como comprobar si se cumplen las condiciones de aplicación de cada una de las técnicas.

También en las técnicas multivariantes, el resultado obtenido es una estimación puntual, por lo que deberá calcularse el correspondiente IC para evaluar su relevancia clínica. Todos los comentarios realizados a propósito de la significación estadística y la relevancia clínica en las pruebas estadísticas bivariantes son aplicables a las multivariantes. La principal diferencia es que en el análisis multivariante se tienen en cuenta simultáneamente las relaciones entre múltiples variables, de forma que una ecuación predictiva mejorará su capacidad de predicción al incorporar más de una variable o la estimación del efecto del factor de estudio estará ajustada, es decir, se habrá controlado el efecto simultáneo de diferentes factores de confusión.

En otras ocasiones no puede diferenciarse entre variables dependientes e independientes, sino que, de acuerdo con el objetivo del estudio, el análisis persigue finalidades diferentes, como la clasificación de variables o individuos aparentemente heterogéneos en grupos homogéneos, según un conjunto de datos (análisis de conglomerados o *cluster analysis*), o la identificación de los factores subyacentes en un conjunto de variables, es decir, la reducción de la dimensionalidad de los datos, combinando un conjunto amplio de variables observadas en unas pocas variables ficticias (combinaciones lineales de las anteriores) que representen casi la misma información que los datos originales (análisis factorial y análisis de componentes principales). Estas técnicas se utilizan con frecuencia en la investigación sobre servicios sanitarios, pero poco en investigación clínica y epidemiológica.

**Ejemplo 27.11.** Supongamos que se desea desarrollar un indicador del estado de salud aplicable a

diferentes regiones. Para ello, se recogen multitud de indicadores sanitarios, demográficos, sociales, etc., de estas regiones, y se aplica una técnica multivariante para reducir todos estos indicadores a unas

pocas variables ficticias, con la extracción de factores, cada uno de los cuales representa una dimensión de la información contenida en la totalidad de los indicadores.

## BIBLIOGRAFÍA

- Altman DG, Bland JM. Parametric vs non-parametric methods for data analysis. *BMJ*. 2009;338:a3167.
- Altman DG. Missing outcomes in randomized trials: addressing the dilemma. *Open Med*. 2009;3(2):e51-3.
- Altman DG, Bland JM. Missing data. *BMJ*. 2007;334:424.
- Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006;332:1080.
- Altman DG, Bland JM. Standard deviations and standard errors. *BMJ*. 2005;331:903.
- Altman D, Bland JM. Confidence intervals illuminate absence of evidence. *BMJ*. 2004;328:1016-7.
- Argimon JM. La ausencia de significación estadística en un ensayo clínico no significa equivalencia terapéutica. *Med Clin (Barc)*. 2002;118:701-3.
- Argimon JM. Intervalos de confianza: algo más que un valor de significación estadística. *Med Clin (Barc)*. 2002;118:382-4.
- Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. 4.ª ed.. Oxford: Blackwell Science; 2002.
- Bayarri MJ, Cobo E. Una oportunidad para Bayes. *Med Clin (Barc)*. 2002;119:252-3.
- Bland JM. The tyranny of power: is there a better way to calculate sample size? *BMJ*. 2009;339:b3985.
- Bland JM, Altman DG. Analysis of continuous data from small samples. *BMJ*. 2009;338:a3166.
- Bland JM, Altman DG. The logrank test. *BMJ*. 2004;328:1073.
- Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*. 2010;303(20):2058-64.
- Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). *BMJ*. 1998;317:1572.
- Bender R, Lange S. Adjusting for multiple testing-when and how? *J Clin Epidemiol*. 2001;54:343-9.
- Bradburn MJ, Clark TG, Love SB, Altman DG. Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. *Br J Cancer*. 2003;89:431-6.
- Clark TG, Bradburn MJ, Love SB, Altman DG. Survival Analysis. Part I. Basic concepts and first analysis. *Br J Cancer*. 2003;89:232-8.
- Cobo E, Muñoz P, González JA, editors. Estadística para no estadísticos: bases para interpretar artículos científicos. 1ª ed. Barcelona: Elsevier; 2007.
- Cobo E. Análisis multivariante en investigación biomédica: criterios para la inclusión de variables. *Med Clin (Barc)*. 2002;119:230-7.
- Cohen HW. P values: use and misuse in medical literature. *Am J Hypertens*. 2011;24:18-23.
- Doll H, Carney S. Introduction to biostatistics: Part 3. Statistical approaches to uncertainty: P values and confidence intervals unpacked. *ACP J Club*. 2006;144:A8-9.
- Katz MH. Multivariable analysis: a primer for readers of medical research. *Ann Intern Med*. 2003;138:644-50.
- Moss M, Wellman DA, Cotsonis GA. An appraisal of multivariable logistic models in the pulmonary and critical care literature. *Chest*. 2003;123:923-8.
- Nüesch E, Trelle S, Reichenbach S, Rutjes AW, Bürgi E, Scherer M, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. *BMJ*. 2009;339:b3244.
- Perneger TV. What's wrong with Bonferroni adjustments. *BMJ*. 1998;316:1236.
- Pocock SJ, Ware JH. Translating statistical findings into plain English. *Lancet*. 2009;373:1926-8.
- Pocock SJ, Travison TG, Wruck LM. How to interpret figures in reports of clinical trials. *BMJ*. 2008;336:1166-9.
- Pocock SJ. The simplest statistical test: how to check for a difference between treatments. *BMJ*. 2006;332:1256-8.
- Royston P, Altman DG. Visualizing and assessing discrimination in the logistic regression model. *Stat Med*. 2010;29(24):2508-20.
- Stone GW, Pocock SJ. Randomized trials, statistics, and clinical inference. *J Am Coll Cardiol*. 2010;55:428-31.
- Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. *Lancet*. 2005;365:1657-61.
- Schulz KF, Grimes DA. Multiplicity in randomised trials I: endpoints and treatments. *Lancet*. 2005;365:1591-5.
- Streiner DL, Norman GR. Correction for multiple testing: is there a resolution? *Chest*. 2011;140:16-8.
- Vedula SS, Altman DG. Effect size estimation as an essential component of statistical analysis. *Arch Surg*. 2010;145:401-2.
- White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ*. 2011;342:d40.

## Potencia de un estudio

Los estudios cuyos resultados no son estadísticamente significativos suelen denominarse *estudios negativos*. En muchas ocasiones se interpretan erróneamente como indicativos de que no existe asociación entre el factor de estudio y la respuesta. Sin embargo, la ausencia de significación estadística no implica necesariamente que no exista asociación, ya que puede haberse producido algún sesgo que la haya enmascarado o haberse estudiado un número insuficiente de sujetos (es decir, tener una baja potencia estadística), de manera que el estudio ha sido incapaz de detectar la asociación (es decir, podría corresponder a un resultado negativo falso). La probabilidad de cometer el error de no detectar como estadísticamente significativa una diferencia o asociación que existe en la realidad se conoce como  $\beta$ . Su complementario  $1 - \beta$  corresponde a la *potencia estadística*, que cuantifica, por tanto, la capacidad de un estudio para detectar como estadísticamente significativa una determinada diferencia o asociación que existe en la realidad.

Como se ha comentado en el capítulo anterior, una prueba estadística calcula, a partir de los resultados observados en un estudio, un valor de  $p$  que indica la probabilidad de rechazar la hipótesis nula ( $H_0$ , ausencia de asociación) cuando es verdadera. El valor de significación estadística  $\alpha$  está fijado por el investigador (habitualmente en el 5%). En cambio, el riesgo  $\beta$  es la probabilidad de no rechazar la  $H_0$  cuando la hipótesis alternativa ( $H_a$ ) es verdadera, por lo que depende del verdadero pero desconocido valor de  $H_a$  (tabla 28.1). En realidad, lo único que se puede demostrar por las observaciones es que una hipótesis es falsa, pero no que sea verdadera, sino que, como máximo, puede decirse que no ha sido contradicha por la experiencia. Se comprende, entonces, que sea prioritario controlar el riesgo de

error: aplicar una prueba a una hipótesis es exigir una evidencia fuerte en favor suyo, y no contentarse con una observación explicable por las fluctuaciones del muestreo.

Como la probabilidad de cometer un error  $\beta$  es diferente según la magnitud de la diferencia que se desea ser capaz de descartar, es importante que se tenga en cuenta este hecho en la determinación del tamaño de la muestra necesario, para intentar garantizar que el estudio tiene la capacidad (potencia) suficiente para detectar una determinada diferencia (si existe).

**Ejemplo 28.1.** En un artículo ya clásico, Freiman et al (1978) revisaron 71 estudios cuyo resultado era negativo, publicados entre 1960 y 1977 en cuatro prestigiosas revistas. Según sus cálculos, la mitad de estos estudios tenía una potencia menor del 26% para encontrar una diferencia relativa del 50%, o menor del 60% si la diferencia que se consideraba importante era del 25%.

**Ejemplo 28.2.** En un estudio más reciente, Dimick et al (2001) evaluaron los 90 ensayos clínicos con resultados negativos publicados en tres revistas de cirugía entre 1988 y 1998. En 55 de ellos (61%) podría considerarse que existía una potencia estadística insuficiente, y tan sólo 22 casos (24%) tenían una potencia superior al 80% para detectar un efecto terapéutico del 50%.

**Ejemplo 28.3.** Bedard et al (2007) revisaron los ensayos clínicos en fase III con resultados negativos presentados en los congresos anuales de la American Society of Clinical Oncology (ASCO) entre 1995 y 2003. De los 423 ensayos analizados, tan sólo 45 (10,6%), 138 (32,6%) y 233 (55,1%) tenían el tamaño de la muestra suficiente (potencia  $\geq 80\%$ ) para

**Tabla 28.1** Errores en una prueba estadística que contrasta la hipótesis de que OR = 1 en función de la hipótesis alternativa considerada

Realidad	Valor de la OR	Conclusión de la prueba	
		Rechazo de $H_0$	No rechazo de $H_0$
$H_0$ verdadera	$OR_0 = 1$	$\alpha$	$1 - \alpha$
$H_a$ verdadera	$OR_a$	$1 - \beta$	$\beta$
	$OR'a$	$1 - \beta'$	$\beta'$

Ho: hipótesis nula.  
Ha: hipótesis alternativa.

detectar efectos de pequeña (*odds ratio* [OR]  $\geq 1,3$ ), mediana (OR  $\geq 1,5$ ) y gran magnitud (OR  $\geq 2,0$ ), respectivamente. Solamente 35 estudios negativos (7,1%) indicaban un motivo para justificar su tamaño insuficiente.

## POTENCIA ESTADÍSTICA

La potencia de un estudio depende de cuatro elementos: la magnitud de la asociación o diferencia existente entre los grupos considerada de interés, la variabilidad de la variable de respuesta, el valor de significación estadística y el número de sujetos estudiado.

La potencia depende de la magnitud de la asociación o la diferencia entre los grupos que se considera de interés o relevancia (mínima diferencia de importancia clínica), es decir, de la hipótesis alternativa  $H_a$ . La potencia es tanto mayor cuanto más se aleja  $H_a$  de  $H_0$ , es decir, cuanto mayor es la diferencia o el efecto que se desea ser capaz de detectar. El aumento de la potencia es mayor de lo que podría parecer intuitivamente.

**Ejemplo 28.4.** Un estudio que compara dos grupos de 200 sujetos cada uno, en el que la proporción observada en el grupo control es del 10%, tiene una potencia del 33% para detectar un riesgo relativo (RR) de 1,5 ( $H_a$ : RR = 1,5), del 81% para detectar un RR de 2 ( $H_a$ : RR = 2) y de hasta el 98% para detectar un RR de 2,5 ( $H_a$ : RR = 2,5).

La potencia también depende de la variabilidad del parámetro de interés. Cuando la variable de respuesta se mide en una escala cuantitativa, puede deducirse intuitivamente que, cuanto mayor sea su variabilidad (mayor variancia), más difícil será poder descartar que determinadas diferencias puedan deberse a las fluctuaciones del muestreo y, por tanto,

menor será la potencia del estudio. Por ello, al diseñar un estudio, las medidas dirigidas a reducir dicha variabilidad (selección de poblaciones homogéneas, reducción de los errores de medida, etc.) aumentarán la potencia del estudio. Lo mismo ocurre cuando la variable se mide en una escala cualitativa. En este caso, la potencia dependerá del valor de la variable de respuesta en el grupo control. Cuando este valor sea muy pequeño, cualquier aumento observado en el grupo de estudio será fácil de detectar.

**Ejemplo 28.5.** Supongamos un ensayo clínico que compara un nuevo fármaco con otro estándar, y que la variable de respuesta es el porcentaje de curaciones. Si el fármaco estándar produce un porcentaje muy bajo de curaciones, cualquier mejoría observada con el nuevo fármaco será fácil de detectar. Si el grupo control tiene, por ejemplo, un porcentaje de éxitos ( $P_c$ ) del 10% y el grupo de estudio ( $P_e$ ) del 20%, esta diferencia absoluta del 10% será más fácil de detectar que en el supuesto de que  $P_c$  fuera del 45% y  $P_e$  del 55%. En el primer caso, la diferencia relativa es del 100% ( $P_e$  es el doble de  $P_c$ ), mientras que en el segundo es sólo del 22%.

El valor de significación estadística prefijado influye directamente sobre la potencia: si se disminuye  $\alpha$  con el fin de reducir el riesgo de llegar a una conclusión falsamente positiva, se aumenta la probabilidad de obtener un resultado falsamente negativo (se disminuye la potencia estadística). Aunque este comentario parece poco relevante porque casi siempre se fija el valor de  $\alpha$  en el 5%, es importante cuando debe decidirse entre el carácter unilateral o bilateral de la prueba. Sin embargo, aunque la prueba unilateral es más potente que la bilateral, la contrapartida es que sólo se presta atención a un sentido de la comparación, por lo que las desviaciones en la dirección contraria conducirán a un resultado estadísticamente no significativo (no rechazo de  $H_0$ ).

Finalmente, cuanto mayor sea el número de participantes en un estudio, mayor será su potencia estadística. Por ello, los estudios con una muestra muy pequeña son los que con más probabilidad pueden conducir a conclusiones falsamente negativas.

## CÁLCULO DE LA POTENCIA ESTADÍSTICA

Por tanto, una vez finalizado un estudio, puede determinarse su potencia estadística a partir de los resultados observados, del número de sujetos realmente analizados y del valor de la significación estadística.

Los métodos de cálculo habituales son aproximados, ya que utilizan valores de parámetros que, de hecho, son desconocidos (p. ej., la variancia cuando se trata de variables cuantitativas, o el valor del porcentaje, si la hipótesis alternativa es cierta, cuando se trata de variables cualitativas), y no tienen en cuenta que la mayoría de los análisis estadísticos que se llevarán a cabo usarán métodos de ajuste o modelos multivariantes.

A continuación se presenta un ejemplo de cálculo de la potencia estadística en un estudio que utiliza una variable de respuesta cualitativa mediante la determinación del valor  $Z_\beta$  (cuadro 28.1). Para determinar la probabilidad de cometer un error  $\beta$  asociada al valor  $Z_\beta$ , puede utilizarse una tabla de la distribución normal (la tabla 28.2 presenta los valores más frecuentes).

**Ejemplo 28.6.** Supongamos un ensayo clínico aleatorio en el que se han estudiado 50 sujetos en cada grupo, y en el que se ha observado un porcentaje de eficacia en el grupo control del 25% ( $p_0 = 0,25$ ). Si se considera como diferencia clínicamente importante que el porcentaje de eficacia del nuevo tratamiento sea como mínimo del 35% ( $p_1 = 0,35$ ), la potencia del ensayo para detectar esta diferencia se calcula del siguiente modo:

$$Z_\beta = \sqrt{\frac{50 \times (0,35 - 0,25)^2}{2 \times 0,3 \times 0,7}} - 1,96 = -0,87$$

(cabe recordar que  $p = [0,25 + 0,35]/2 = 0,30$ , por lo que  $q = 1 - 0,30 = 0,70$ ).

A partir de la tabla 28.2 puede apreciarse que la potencia estadística se sitúa entre un 18 y un 21%. Si se busca el valor de  $Z_\beta = -0,87$  en las tablas de la distribución normal, se obtiene que la probabilidad de cometer un error  $\beta$  es de 0,81, es decir, que la potencia del estudio para detectar esta diferencia es sólo de 0,19 (19%).

**Cuadro 28.1 Fórmula para el cálculo de la potencia estadística cuando la variable de respuesta es cualitativa**

$$Z_\beta = \sqrt{\frac{n(p_1 - p_0)^2}{2_{pq}}} - Z_{\alpha/2}$$

$Z_\beta$ : es el valor de la distribución normal Z para un valor  $\beta$  determinado (unilateral).

$n$ : es el número de sujetos estudiados en cada grupo (se asume que es el mismo en ambos grupos).

$p_0$ : es el porcentaje de eficacia observado en el grupo control.

$p_1$ : es el porcentaje de eficacia del grupo de intervención que se desearía detectar (así pues,  $p_1 - p_0$  sería la diferencia de relevancia clínica que se desearía detectar).

$p$ :  $(p_0 + p_1)/2$ .

$q$ :  $= 1 - p$ .

$Z_{\alpha/2}$ : es el valor de la distribución normal Z para el valor  $\alpha$  bilateral fijado (habitualmente,  $\alpha = 0,05$  y  $Z_{\alpha/2} = 1,96$ ).

Existe una fórmula diferente para cada una de las situaciones posibles. De todas formas, puede ser útil un método sencillo que permite determinar, si se conoce el número de sujetos que era necesario para detectar una determinada diferencia con la potencia deseada al inicio del estudio, la diferencia que ha sido capaz de detectar con el número de sujetos realmente estudiado y la potencia deseada:

$$d_2 = d_1 \sqrt{\frac{n_1}{n_2}}$$

donde  $d_2$  es la diferencia entre ambos grupos que podría detectarse con el número de sujetos estudiado;  $d_1$  es la diferencia considerada de relevancia clínica, a partir de la cual se determinó el número de sujetos necesario;  $n_1$  es el número de sujetos que se determinó que era necesario para detectar la diferencia  $d_1$ ; y  $n_2$  es el número de sujetos realmente estudiado.

**Ejemplo 28.7.** Supongamos que se ha realizado un ensayo clínico para evaluar la eficacia de un nuevo fármaco. La eficacia del tratamiento utilizado como control es del 25% ( $p_0 = 0,25$ ). El estudio se diseñó de forma que, para detectar una diferencia entre los tratamientos del 10% ( $p_1 - p_0 = 0,1$ ) con un valor de significación bilateral del 5%, tuviera una potencia del 90%, lo que significaba que eran necesarios 439 sujetos por grupo. Sin embargo, al realizar el estudio, sólo se encontraron 600 sujetos disponibles,

**Tabla 28.2 Tabla de conversión de valores de  $Z_\beta$  a potencia estadística**

$Z_\beta$	Potencia	$Z_\beta$	Potencia
-2,5	0,006	0,0	0,500
-2,4	0,008	0,1	0,540
-2,3	0,011	0,2	0,579
-2,2	0,014	0,3	0,618
-2,1	0,018	0,4	0,655
-2,0	0,023	0,5	0,691
-1,9	0,029	0,6	0,726
-1,8	0,036	0,7	0,758
-1,7	0,045	0,8	0,788
-1,6	0,055	0,9	0,816
-1,5	0,067	1,0	0,841
-1,4	0,081	1,1	0,864
-1,3	0,097	1,2	0,885
-1,2	0,115	1,3	0,903
-1,1	0,136	1,4	0,919
-1,0	0,159	1,5	0,933
-0,9	0,184	1,6	0,945
-0,8	0,212	1,7	0,955
-0,7	0,242	1,8	0,964
-0,6	0,274	1,9	0,971
-0,5	0,309	2,0	0,977
-0,4	0,345	2,1	0,982
-0,3	0,382	2,2	0,986
-0,2	0,421	2,3	0,989
-0,1	0,460	2,4	0,992
0,0	0,500	2,5	0,994

300 en cada grupo. ¿Qué diferencia ha sido capaz de detectar este estudio con el número de sujetos estudiado con la potencia deseada del 90%?

$$d_2 = 0,1 \quad \frac{439}{300} = 0,12$$

Por lo tanto, al haber estudiado 300 sujetos por grupo, se tenía una potencia del 90% para detectar una diferencia mínima del 12%; es decir, dado que la eficacia del tratamiento de control es del 25%, podrá detectarse un aumento de la eficacia con el nuevo tratamiento a partir del 37%.

## POTENCIA Y DISEÑO DEL ESTUDIO

Lo verdaderamente importante no es determinar la potencia estadística una vez finalizado el estudio, sino diseñarlo de manera que tenga la mayor potencia posible y así aumentar su capacidad para detectar el resultado de interés. Numerosos aspectos del diseño del estudio están relacionados con su potencia. La mayoría de ellos son interdependientes y, a menudo, resulta imposible optimizarlos de forma simultánea.

## Definición de la población de estudio

Si bien la elección de la población es obligada por la misma naturaleza del estudio y de la hipótesis que se desea contrastar, es conveniente tener en cuenta algunos aspectos que pueden aumentar la potencia.

En primer lugar, ya que la potencia depende de la variabilidad de la variable de respuesta, la selección de una población homogénea tiende a aumentar la potencia.

Por otro lado, la potencia de un estudio es tanto mayor cuanto más se aproxima al 50% la proporción de sujetos expuestos (en los estudios de casos y controles) o la frecuencia de la enfermedad (en los estudios de cohortes y ensayos clínicos). Pero, en la práctica, raramente se alcanzan estos valores, por lo que se trata de definir una población de estudio en la que estos porcentajes sean lo más elevados posible.

Dado que la potencia también depende de la magnitud de la asociación considerada ( $H_a$ ), es conveniente seleccionar una población de estudio en la que sea probable que el efecto sea de mayor magnitud (p. ej., utilizando a personas de edad avanzada o niños asmáticos para evaluar el riesgo respiratorio asociado a la contaminación atmosférica).

El empleo de criterios de selección amplios con la finalidad de conseguir un mayor número de casos no necesariamente aumenta la potencia del estudio, ya que puede introducir una mayor variabilidad y diluir la asociación, al incluir casos en los que ésta sea menor o en los que ésta no esté asociada con la exposición. Por ejemplo, en ocasiones, dado el previsible reducido número de casos, los investigadores están tentados de añadir categorías diferentes de la enfermedad. Sin embargo, esta práctica puede disminuir la potencia y conducir a un sesgo en la estimación del efecto, ya que aumenta la heterogeneidad de la muestra y determinados efectos pueden quedar diluidos.

**Ejemplo 28.8.** Supongamos un estudio de cohortes que investiga si la administración de un fármaco durante el embarazo aumenta la incidencia de una determinada malformación congénita. Dado que ésta sólo aparece en un número muy reducido de casos, los investigadores deciden incluir todas las malformaciones, con independencia de su tipo, en una variable de respuesta única. Como no existe un teratógeno único que cause todas las malformaciones, el problema del análisis puede ser agrupar malformaciones no relacionadas con el medicamento de estudio con otras que sí lo están.

Las estrategias para prevenir fenómenos de confusión (emparejamiento y estratificación) también

son útiles para aumentar la potencia estadística, al reducir la variabilidad del parámetro de estudio.

## Medición del factor de estudio y la variable de respuesta

Los errores de medida conducen a sesgos de clasificación, una de cuyas consecuencias es una pérdida de la potencia causada por una dilución de la asociación. Por otro lado, no debe olvidarse que la propia elección de la variable de medida de la exposición al factor de estudio y de la respuesta tiene una gran influencia sobre la potencia, ya que de ella depende, en gran parte, la capacidad para detectar una diferencia o una asociación. Por ello es fundamental asegurar que las variables estén perfectamente definidas al inicio del estudio y que su medida sea lo más exacta posible.

## Tamaño de la muestra

El número de sujetos incluidos en un estudio influye directamente sobre su potencia. En general, la máxima potencia se alcanza cuando los grupos tienen el mismo tamaño. Sin embargo, en ocasiones, el número de sujetos accesibles está limitado en uno de los grupos (habitualmente los casos en los estudios de casos y controles, y los expuestos en los de cohortes), de forma que si se elige el mismo número en el otro grupo, la potencia será muy reducida. En estas circunstancias se puede incrementar el número de sujetos del otro grupo de comparación para aumentar la potencia, aunque esta ganancia es casi nula cuando la proporción es superior a 4:1.

La ausencia de información sobre algunos sujetos puede comportar una disminución de la potencia al reducirse el número de individuos útiles para el análisis, especialmente cuando se utilizan técnicas multivariantes, ya que éstas sólo tienen en cuenta a los sujetos de los que se dispone de información en todas las variables implicadas.

## POTENCIA Y ANÁLISIS ESTADÍSTICO

El investigador también puede influir sobre algunos aspectos del análisis de los datos con la finalidad de aumentar la potencia estadística.

## Características de la prueba estadística

Una prueba unilateral es más potente que una bilateral. Supongamos que se desea evaluar el exceso

de riesgo asociado a una exposición. Para utilizar una prueba unilateral ( $H_a$ : OR > 1), no sólo debe ser ésta la hipótesis de interés, sino que deben existir fuertes argumentos para excluir la posibilidad de que el verdadero valor de OR sea inferior a 1. De hecho, si se observa un valor de OR inferior a 1, el resultado se considerará estadísticamente no significativo, es decir, será atribuido a las fluctuaciones del muestreo alrededor del valor OR = 1, y no se considerará el reflejo de un valor de OR en la población inferior a 1.

### Elección y definición de la variable

Si una variable se mide en una escala cuantitativa, en general es preferible no categorizarla para no perder información. Además, la elección de los intervalos que definen las categorías puede conducir a un resultado significativo, que no aparecería de otro modo. Por ello, es importante que las diferentes alternativas sean discutidas con independencia de los resultados observados.

El uso de variables combinadas para medir un efecto permite aumentar la potencia estadística, aunque debe tenerse en cuenta que su utilización puede diluir determinadas diferencias o asociaciones que afecten a algunos de sus componentes.

En algunas situaciones, la definición de un período de exposición de interés puede influir sobre la potencia del estudio.

**Ejemplo 28.9.** El riesgo de cáncer asociado a la exposición ambiental a determinados agentes químicos suele manifestarse tras períodos de inducción muy largos, entre 10 y 40 años. Por ello, cuando se detecta la aparición de uno de estos cánceres en un individuo, las exposiciones muy recientes (p. ej., en los últimos 10 años) o muy antiguas (p. ej., hace más de 40 años) no deberían incluirse en el cálculo de la exposición acumulada, el cual debería limitarse a una ventana temporal (en el ejemplo, 10-40 años antes de la aparición del cáncer). La inclusión de exposiciones ocurridas fuera de este intervalo podría

producir una dilución de la asociación y, por tanto, una pérdida de la potencia.

### Ajuste y sobreajuste

El *ajuste* tiene como finalidad controlar los factores de confusión, pero también permite disminuir la variabilidad del parámetro estudiado, lo que conduce a un aumento de la potencia.

Se habla de *sobreajuste* cuando, al ajustar por una variable, se reduce la variabilidad del factor de estudio y, por tanto, la potencia de la prueba, sin que ello signifique el control de un factor de confusión. Este fenómeno puede aparecer cuando se controla por una variable muy correlacionada con el factor de estudio, ya que puede ser difícil distinguir los efectos de ambos, lo que conduce a una estimación imprecisa de la asociación. Esta pérdida de precisión en la estimación debida al sobreajuste también puede aparecer cuando se controla por variables que no son verdaderos factores de confusión.

## POTENCIA E INTERPRETACIÓN DE LOS RESULTADOS

Al diseñar un estudio, los investigadores fijan la potencia estadística deseada al determinar el número mínimo de sujetos necesario para encontrar una diferencia que se considera clínicamente útil, con unos márgenes de error  $\alpha$  y  $\beta$  prefijados. No obstante, es frecuente que las cosas no salgan como se habían planeado, por lo que, si el resultado no es estadísticamente significativo, es conveniente calcular la potencia real del estudio para haber detectado dicha diferencia. Si es elevada, lo más probable es que el resultado refleje que, en realidad, la diferencia no existe o es inferior a la magnitud que se deseaba descartar. Si es reducida, el resultado es poco informativo y debe interpretarse como que las evidencias obtenidas son insuficientes para llegar a una conclusión sobre la existencia o la ausencia de la diferencia.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

Bedard PL, Krzyzanowska MK, Pintilie M, Tannock IF. Statistical power of negative randomized controlled trials presented at American Society for Clinical Oncology annual meetings. *J Clin Oncol*. 2007;25:3482-7.

Dimick JB, Diener-West M, Lipsett PA. Negative results of randomized clinical trials published in the surgical literature: equivalency or error? *Arch Surg*. 2001;136:796-800.  
Freiman JA, Chalmers TC, Smith H, Kuebler RR. The importance

of beta, type II error and sample size in the design and interpretation of the randomized control trial. *N Engl J Med*. 1978;299:690-4.

## BIBLIOGRAFÍA

- Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ*. 1995;311:485.
- Bhardwaj SS, Camacho F, Derrow A, Fleischer Jr AB, Feldman SR. Statistical significance and clinical relevance: the importance of power in clinical trials in dermatology. *Arch Dermatol*. 2004;140:1520-3.
- Fitzner K, Heckinger E. Sample size calculation and power analysis: a quick review. *Diabetes Educ*. 2010;36:701-7.
- Galbraith S, Marschner IC. Guidelines for the design of clinical trials with longitudinal outcomes. *Control Clin Trials*. 2002;23:257-73.
- Godwin M. Hypothesis: the research page. Part 3: Power, sample size, and clinical significance. *Can Fam Physician*. 2001;47:1441-3.
- Karlsson J, Engebretsen L, Dainty K. ISAKOS Scientific Committee. Considerations on sample size and power calculations in randomized clinical trials. *Arthroscopy*. 2003;19:997-9.
- Keen HI, Pile K, Hill CL. The prevalence of underpowered randomized clinical trials in rheumatology. *J Rheumatol*. 2005;32:2083-8.
- Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA*. 1994;272:122-4.
- Okumura Y, Sakamoto S. Statistical power and effect sizes of depression research in Japan. *Psychiatry Clin Neurosci*. 2011;65:356-64.
- Sahai H, Khurshid A. Formulae and tables for the determination of sample sizes and power in clinical trials for testing differences in proportions for the two-sample design: a review. *Stat Med*. 1996;15:1-21.
- Tsang R, Colley L, Lynd LD. Inadequate statistical power to detect clinically significant differences in adverse event rates in randomized controlled trials. *J Clin Epidemiol*. 2009;62:609-16.
- van Kleef M, Kessels AG. Underpowered clinical trials: time for a change. *Pain*. 2009;145:265-6.

## Sesgos

Los sesgos se definen como errores sistemáticos que se introducen en un estudio y que dan como resultado una estimación incorrecta (sesgada) del parámetro o del efecto de interés. Por lo tanto, la ausencia de errores sistemáticos, o sesgos, confiere validez interna a un estudio.

A grandes rasgos, los sesgos se pueden deber a la forma en que se seleccionan los sujetos de estudio (*sesgos de selección*) o a la manera en que se recoge la información (*sesgos de información*).

Una vez que aparece un sesgo no puede eliminarse en el análisis, a diferencia de lo que ocurre con los factores de confusión. Por esta razón, es fundamental prevenir su aparición en la fase de diseño. Sin embargo, en la fase de interpretación es bueno volverse a formular las mismas preguntas que se hicieron en los inicios del estudio, cuando se escogió la población y las variables que iban a ser medidas (tabla 29.1), lo que ayudará a detectar posibles errores que hayan pasado desapercibidos mientras el estudio se llevaba a cabo y que pueden ser útiles para discutir los resultados.

Cuando se interpreta un estudio, debe explorarse la posible presencia de sesgos como una explicación alternativa a los resultados observados. Para ello hay que tener en cuenta el tipo de estudio y las características específicas de su diseño, si los sesgos son de selección y/o información, y, en este último supuesto, si se trata de un error diferencial o no diferencial. No interesa sólo saber si ha introducido un sesgo, sino también conocer su posible efecto sobre los resultados observados.

### SESGOS DE SELECCIÓN

En los estudios transversales, la forma en que se selecciona la muestra de sujetos o la presencia de no respuestas pueden introducir sesgos de selección.

En los estudios longitudinales analíticos, los sesgos de selección se presentan, principalmente, en los estudios de casos y controles, y en los estudios retrospectivos de cohortes, donde tanto la exposición (factor de estudio) como la enfermedad (respuesta) han ocurrido en el momento de iniciarse el estudio. En los estudios de casos y controles, los sesgos de selección pueden ocurrir cuando la identificación de los individuos para su inclusión en el estudio se basa en criterios distintos según se trate de casos o de controles y, además, estos criterios están relacionados con la exposición en estudio. Análogamente, en los estudios de cohortes, si la clasificación de los individuos en «expuestos» y «no expuestos» depende del desarrollo de la enfermedad de interés, también se producirá un sesgo de selección.

Por el contrario, es menos frecuente que se produzcan en los ensayos clínicos aleatorios, ya que el único factor que interviene en la asignación de los individuos a los grupos de estudio es el azar. De todos modos, en los estudios prospectivos, ya sean de cohortes o ensayos clínicos, las pérdidas durante el seguimiento pueden dar lugar a un sesgo de selección.

Los sesgos de selección pueden ocurrir en numerosas circunstancias, relacionadas con la forma de identificar y seleccionar a los individuos para un estudio (cuadro 29.1).

### Selección del grupo control

Los estudios analíticos se basan en que el grupo de estudio y el de referencia son comparables por las variables pronósticas o predictoras de la respuesta, de forma que sólo difieren en la presencia o no del factor de estudio (principio de comparabilidad). En los ensayos clínicos aleatorios, la asignación aleatoria de los sujetos a los grupos de estudio tiende a asegurar esta

**Tabla 29.1** Preguntas que se han de formular para evaluar la presencia de un sesgo

Pregunta	Tipo de sesgo si la respuesta es negativa	Observaciones
¿Todos los participantes tienen la misma probabilidad de haber recibido la intervención o haber estado expuestos al factor de estudio?	Sesgo de selección	En los ensayos clínicos se previene con la asignación al azar En los estudios observacionales se previene siguiendo unos criterios de selección y exclusión muy estrictos
¿Los individuos que se pierden tienen la misma probabilidad de desarrollar la respuesta que los que finalizan el estudio?	Sesgo de selección que limita la comparabilidad final del estudio	Hay que conocer los motivos y estudiar una muestra de los que se pierden para poder contestar a la pregunta y poder interpretar los resultados
¿La detección de la respuesta se ha hecho con las mismas pruebas y con la misma intensidad en todos los participantes?	Sesgo de información	Emplear técnicas de enmascaramiento Dar resultados falsos negativos (atenuar la asociación)
¿La medición de la variable de respuesta es sensible y específica?	Sesgo de información	En algunas ocasiones puede estimarse su magnitud o corregirse una vez finalizado el estudio

**Cuadro 29.1** Situaciones en las que puede haberse introducido un sesgo de selección en un estudio

- Selección del grupo control
- Pérdidas de seguimiento
- Supervivencia selectiva
- Muestra no representativa de la población de estudio
- Sesgo de detección
- Participación de voluntarios o efecto de autoselección

comparabilidad al inicio del estudio, mientras que este principio es más difícil de asegurar en los estudios observacionales, especialmente en los diseños de casos y controles y en los retrospectivos de cohortes.

**Ejemplo 29.1.** Consideremos un estudio retrospectivo de cohortes que trataba de analizar el efecto de un programa de educación para diabéticos sobre el control metabólico de la enfermedad. Para ello los investigadores incluyeron en el grupo de estudio 40 individuos que cumplían los siguientes requisitos: *a) haber sido tratados por primera vez en su centro entre 1994 y 1996; b) haber cumplido los cinco apartados de los que constaba el programa de educación, y c) haber cumplido con las visitas de seguimiento.* El grupo control se formó contactando telefónicamente

de manera consecutiva con aquellas personas de su lista de pacientes que cumplían las siguientes condiciones: *a) haber sido tratados por primera vez en su centro entre 1994 y 1996; b) haber rechazado el programa de educación, y c) haber acudido o no a las visitas de seguimiento.*

En este estudio se observó que los individuos que habían recibido el programa poseían un mejor control metabólico que la cohorte de referencia. Sin embargo, no se podía asegurar que ello fuera debido a haber recibido el programa de educación. Al revisar los criterios de selección de ambos grupos, puede observarse que son diferentes, ya que el grupo control está formado por sujetos que habían rechazado el programa de educación y que podían no haber acudido a las visitas de seguimiento. Así pues, estos dos hechos parecen indicar que el grupo de referencia podría estar formado por individuos menos motivados por su enfermedad que los del grupo de estudio, de forma que podría ser esta diferente motivación la que les hubiera conducido a un peor control de la diabetes, independientemente del efecto del programa de educación. Este ejemplo pone de manifiesto la dificultad que existe en los estudios retrospectivos para evaluar determinadas prácticas preventivas o exposiciones, debido a la introducción de sesgos de selección.

La elección de un grupo de comparación adecuado que evite los sesgos de selección es particularmente

difícil en los estudios de casos y controles. La selección de los sujetos del grupo control ha de ser independiente de la historia de exposición, y debe excluir aquellos individuos que tengan enfermedades relacionadas de forma positiva o negativa con la exposición. Por otro lado, también se recomienda incluir pacientes con diferentes patologías como grupo control, y comprobar que las frecuencias de exposición son similares entre ellas. La finalidad de esta estrategia es que, si alguna de las patologías seleccionada está asociada de forma positiva o negativa con la exposición, su influencia dentro del grupo control sea pequeña.

### Pérdidas de seguimiento

En los estudios prospectivos, una de las principales causas potenciales de sesgos de selección son las pérdidas de individuos, o abandonos, producidas durante el período de seguimiento. Cuando las personas que se pierden durante el seguimiento difieren de las que no abandonan por variables relacionadas con el factor de estudio y la variable de respuesta, se producirá un sesgo en los resultados. La introducción de este sesgo dependerá de los motivos de la pérdidas, más que de su número.

**Ejemplo 29.2.** Supongamos un estudio de cohortes en el que se sigue un grupo de sujetos fumadores y otro de no fumadores, para comparar en ellos la incidencia de cáncer de pulmón. Durante el período de estudio es muy probable que se produzcan pérdidas de seguimiento con mayor frecuencia en la cohorte de fumadores debido a la aparición de problemas de salud relacionados con el consumo de tabaco (p. ej., un mayor número de defunciones por infarto de miocardio). Si no se tiene en cuenta este hecho, se producirá un sesgo en la estimación del riesgo de cáncer de pulmón asociado al consumo de tabaco, ya que probablemente los sujetos más expuestos (más fumadores) habrán presentado otros problemas de salud relacionados, de forma que la incidencia de cáncer se estimará en un grupo de sujetos que estarán menos expuestos que la cohorte inicial.

### Supervivencia selectiva

El sesgo de selección debido a la supervivencia selectiva se produce en los estudios en que se incluyen casos prevalentes, es decir, sujetos que han sobrevivido a las primeras fases de su enfermedad. Si la letalidad de una enfermedad depende de la exposición (p. ej., si la cardiopatía isquémica [CI] es más grave en los sujetos que presentan más factores de riesgo), los casos prevalentes no serán representativos de todos los casos, sino que presentarán menores niveles de exposición, lo que puede conducir a una estimación sesgada de la asociación. Además, los casos prevalentes pueden haber cambiado sus hábitos en relación con la exposición (p. ej., haber dejado de fumar tras un infarto de miocardio), hecho que también conduciría a un sesgo en los resultados.

**Ejemplo 29.3.** Supongamos un estudio de cohortes en el que se incluyen 1.000 individuos expuestos a un factor de riesgo (p. ej., hipertensión arterial) y 1.000 no expuestos. Ambos grupos son seguidos durante 10 años con el fin de comparar las tasas de incidencia de CI observadas en cada uno de ellos. Los resultados a los 10 años se presentan en la tabla 29.2. El riesgo de CI sería tres veces mayor en los expuestos que en los no expuestos (300 casos de CI entre los expuestos y sólo 100 entre los no expuestos).

Supongamos ahora que, en lugar de haberse llevado a cabo este estudio, se hubiera realizado un estudio de casos y controles 10 años después, seleccionando los casos prevalentes (supervivientes). Por lo tanto, se incluirían como casos los sujetos vivos con CI (primera columna de resultados de la tabla), y como controles, los vivos sin CI (tercera columna de resultados de la tabla). En este estudio hipotético, la estimación del riesgo de CI sería  $OR = (50 \times 900) / (700 \times 80) = 0,80$ . Dado que la OR es inferior a 1, este resultado parece sugerir que la hipertensión arterial habría actuado como un factor protector de la CI. Por estos motivos, en los estudios de casos y controles sobre una enfermedad con elevada mortalidad deben utilizarse siempre casos incidentes.

**Tabla 29.2 Resultados de un estudio hipotético en el que se compara la incidencia de cardiopatía isquémica (CI) de dos cohortes de 1.000 sujetos, seguidas durante 10 años**

	Vivos con CI	Muertos por CI	Vivos sin CI	Total
Expuestos	50	250	700	1.000
No expuestos	80	20	900	1.000
Total	130	270	1.600	2.000

## Muestra no representativa de la población de estudio

La selección de una muestra no representativa es una causa frecuente de sesgo de selección en los estudios transversales, aunque también puede presentarse en los estudios analíticos. Se introduce en el proceso de selección de la muestra a partir de la población de estudio.

**Ejemplo 29.4.** Consideremos que se realiza un estudio transversal para estimar la prevalencia de un problema de salud (p. ej., hipercolesterolemia) en una comunidad determinada. Si se selecciona la muestra a partir de los individuos que acuden de forma espontánea a los centros sanitarios públicos por cualquier motivo, probablemente se obtendrá una estimación sesgada de la prevalencia, ya que las personas que consultan no son representativas de la población general, ni las que lo hacen en centros públicos son comparables a las que acuden a centros privados.

En este tipo de estudios debe identificarse claramente la población de estudio y obtener una muestra de ella mediante una técnica de selección que tienda a asegurar su representatividad y prever estrategias para reducir el número de no respuestas.

## Sesgo de detección

El sesgo de detección se produce cuando el procedimiento utilizado para identificar la presencia de la respuesta varía o se modifica según la presencia o la ausencia del factor de estudio. Por ejemplo, en los estudios de casos y controles, este sesgo podría presentarse si los casos fueran individuos en los que la presencia de enfermedad hubiera tenido mayor probabilidad de diagnosticarse por la presencia de la exposición.

**Ejemplo 29.5.** Un ejemplo clásico de este tipo de sesgo es el estudio sobre el riesgo de tromboembolia pulmonar (TP) asociado al uso de anticonceptivos orales (AO). En el primer estudio de casos y controles publicado sobre esta asociación (Sartwell et al, 1969) se discutió que una posible explicación era el hecho de que algunos médicos conocían la hipótesis de trabajo y, en consecuencia, las mujeres que tomaban AO tenían mayor probabilidad de ser ingresadas con el diagnóstico de TP y, además, podían haber sido sometidas a un mayor número de pruebas para diagnosticar o descartar la enfermedad que las mujeres del grupo control. Si esto hubiera sucedido realmente, la observación de una mayor frecuencia de uso de AO en las mujeres con TP podría explicarse, al menos en parte, porque la

hospitalización y la determinación del diagnóstico habrían estado influidas por el uso de AO.

## Participación de voluntarios o efecto de autoselección

Cuando se decide utilizar voluntarios en un estudio, debe tenerse en cuenta que pueden introducir un sesgo, ya que estos sujetos habitualmente son diferentes de los no voluntarios (autoselección). Si la voluntad de participar está relacionada con factores que pueden influir sobre la relación que se estudia, la selección de este tipo de sujetos puede conducir a una estimación sesgada.

**Ejemplo 29.6.** Un ejemplo de este tipo de sesgo es la evaluación del efecto de la vacuna antipoliomielítica de Salk. Se utilizaron dos diseños diferentes para valorar la eficacia de la vacuna (Meier, 1978). En algunas comunidades se asignaron los niños al azar para recibir la vacuna o una inyección de placebo. Sin embargo, otras comunidades rehusaron participar en un ensayo clínico aleatorio, pero estuvieron de acuerdo en que la vacuna se podía administrar a los niños de familias que mostraban su interés en recibirla y que los demás podían servir de control. Al analizar los datos, los investigadores encontraron que las familias que aceptaron voluntariamente la participación de sus hijos tenían un nivel de estudios superior e ingresos económicos mayores que los del grupo control. Aunque estas variables se puedan controlar en el análisis, da idea de que ha existido una cierta selección por múltiples factores, algunos de ellos no medidos, y que pueden sesgar la estimación del efecto, en este caso, la eficacia de la vacuna.

Este proceso de selección puede producirse antes de que los sujetos sean identificados para su inclusión en el estudio. Por ejemplo, es bien conocido que la mortalidad de los trabajadores en activo es menor que la de la población de la misma edad y el mismo sexo en su conjunto. Este efecto del *trabajador sano* se debe presumiblemente a un fenómeno de autoselección, que permite a la gente que goza de buena salud convertirse en (o mantenerse como) trabajador en activo, en tanto que los que permanecen desempleados, retirados, incapacitados o de algún modo fuera de la población trabajadora activa son, en general, un grupo que goza de menos salud.

## SESGOS DE INFORMACIÓN

Se producen cuando la información sobre el factor de estudio o la variable de respuesta es errónea o se recoge de forma sistemáticamente diferente entre

los grupos de estudio. Son causas frecuentes el uso de instrumentos de medida inadecuados o de definiciones poco válidas o precisas de las variables, los errores introducidos por los encuestadores o por el personal que recoge la información, o errores debidos a los propios participantes en el estudio, como el *sesgo de memoria*. En el diseño del estudio debe preverse su aparición, utilizando definiciones e instrumentos de medida válidos y asegurando que el personal que los aplica lo haga de igual forma en los distintos grupos.

Cuando la finalidad de un estudio es la estimación de un parámetro en una muestra de sujetos, la presencia de un sesgo de este tipo conduce a la obtención de un resultado no válido. En los estudios analíticos, la introducción de un sesgo de información conduce a una mala clasificación de los sujetos respecto al factor de estudio o a la variable de respuesta. Las consecuencias son diferentes según si el error de clasificación respecto al estado de expuesto o no expuesto en los estudios de casos y controles (o de enfermo o no enfermo en los estudios de cohortes) es o no independiente del estado de enfermo o no enfermo (o expuesto o no expuesto, en los estudios de cohortes). Cuando los errores de clasificación no son independientes se habla de *mala clasificación diferencial*, y cuando lo son, de *mala clasificación no diferencial*.

### Error de clasificación no diferencial

Cuando la proporción de sujetos clasificados erróneamente es similar en cada uno de los grupos de estudio, se produce un error de clasificación no diferencial. Un ejemplo de esta situación es el empleo de una técnica poco sensible para medir la exposición al factor de riesgo, que se aplica por igual a los individuos de ambos grupos.

Todos los estudios de investigación clínica y epidemiológica tienen siempre algún grado de mala clasificación de la exposición al factor de estudio o de la variable de respuesta, ya que generalmente es imposible utilizar un proceso de medición que tenga a la vez un 100% de sensibilidad y un 100% de especificidad.

Los errores de clasificación no diferenciales tienden a producir, aunque no siempre, una infraestimación de la verdadera diferencia o asociación, es decir, que la estimación estará sesgada hacia la unidad cuando se expresa en medidas relativas, o hacia el cero si se expresa como una diferencia absoluta. Por esta razón, para la validez de un estudio estos errores suelen considerarse una amenaza menor que los errores diferenciales. Sin embargo, pueden ser un problema importante, ya que pueden

explicar algunas discrepancias entre los resultados de diferentes estudios epidemiológicos.

El grado de infraestimación que producen depende no sólo de la sensibilidad y la especificidad del instrumento de medida utilizado, sino también de la prevalencia de la exposición.

**Ejemplo 29.7.** Supongamos un estudio en que se mide la exposición al factor de estudio con un instrumento que tiene una sensibilidad del 90% y una especificidad también del 90%, y en el que la prevalencia de exposición es del 50%. Si la verdadera magnitud de la asociación entre la exposición y la enfermedad fuera una *odds ratio* (OR) = 5, en el estudio se observaría una OR = 3,29. Si la prevalencia de exposición fuera del 1%, el valor de la OR observado en el estudio sería de 1,33.

Este hecho puede ser importante cuando se realiza un ajuste por factores de confusión, ya que, si la prevalencia de la exposición es diferente en cada una de las categorías por las que se estratifica, el efecto del error de medida será diferente en cada una de ellas. En el ejemplo 29.7 puede verse que, si la prevalencia de exposición fuera del 1% en una categoría del factor de confusión y del 50% en la otra, se obtendrían OR de 1,3 y 3,3, respectivamente, lo cual daría la falsa impresión de que existe una modificación de efecto, cuando en realidad la OR es de 5 en ambas situaciones. Por lo tanto, si la variable de confusión se ha medido sin error, pero la exposición (o la respuesta) se ha medido incorrectamente, los resultados deben interpretarse con cautela, ya que pueden ser artefactos producidos por el error de medida. Cuando es la propia variable de confusión la que está sujeta a un error de medida no diferencial, se puede observar la misma falsa apariencia de modificación del efecto del factor de estudio sobre la respuesta.

Otra posible consecuencia del error de medición de una variable de confusión es que, al ajustar la estimación por dicha variable, puede quedar un efecto de confusión residual.

Los ejemplos y comentarios anteriores ilustran los peligros de la utilización de medidas de baja calidad, no sólo a la hora de clasificar a los individuos según su exposición al factor de riesgo o su estado de enfermedad, sino incluso de las variables de confusión por las que se va a ajustar.

En la interpretación de los resultados de un estudio que no ha observado un efecto, debe tenerse en cuenta la posibilidad de que se haya introducido un sesgo de clasificación no diferencial, para evaluar en qué medida podría haber enmascarado un efecto real. Por el contrario, tiene menos importancia en

estudios en los que se ha observado un efecto importante, ya que este sesgo lo que produce es una infraestimación del mismo. Consecuentemente, en términos generales, no es correcto desdeñar un estudio que ha observado un efecto simplemente por el hecho de que exista un error de clasificación no diferencial.

### Error de clasificación diferencial

Los errores de clasificación diferencial pueden ir tanto en la dirección de infraestimar la magnitud de la asociación o de la diferencia, como en la de sobreestimarla.

Un ejemplo de error de clasificación diferencial es el denominado *sesgo de memoria*, que se produce cuando los individuos con un determinado problema de salud recuerdan su historia de exposición de forma distinta a cómo la recuerdan aquellos que no tienen dicho problema. Suele producirse cuando se recoge la información sobre la exposición mediante una entrevista o un cuestionario, y los casos recuerdan sus antecedentes de forma diferente a los controles. También puede producirse cuando los individuos expuestos a algún riesgo (ocupacional, ambiental, etc.) informan sobre el desarrollo de problemas de salud de forma diferente a como lo hacen los que no están expuestos. Este tipo de sesgo es especialmente problemático en los estudios de casos y controles, y en los retrospectivos de cohortes, ya que tanto la exposición como la enfermedad ya han ocurrido, y puede conducir tanto a una sobreestimación como a una infraestimación de la asociación entre la exposición y la enfermedad, dependiendo de si los casos recuerdan su exposición mejor o peor que los controles.

**Ejemplo 29.8.** En un estudio realizado en madres de niños con malformaciones congénitas, se obtuvo información mediante una entrevista personal sobre 8 posibles exposiciones durante el embarazo y se comparó con la información registrada en la historia clínica del obstetra (Werler et al., 1989). Se consiguió la misma información, y de las mismas fuentes, de un grupo de madres de hijos sin malformaciones, que sirvió de control. La información sobre 5 de las 8 exposiciones fue igual de exacta en los casos que en los controles, pero la historia de infección durante el embarazo, de infertilidad tratada o no, y del uso de métodos anticonceptivos durante 2 semanas o más después del último período menstrual, fue mucho más exacta en los casos que en los controles.

Otro tipo de sesgo diferencial es el debido al entrevistador o a la persona que recoge la información,

que aparece cuando existe alguna diferencia sistemática en la forma en que solicita, recoge o interpreta los datos procedentes de los participantes en un estudio, en función del grupo al que pertenecen.

**Ejemplo 29.9.** En un estudio de casos y controles cuyo objetivo era evaluar la hipótesis de que la endometriosis predispone a la infertilidad, se revisaron las historias clínicas de 100 mujeres en las que se había practicado una laparoscopia a causa de una infertilidad (Strathy et al., 1982). En 21 de ellas se diagnosticó endometriosis. Como grupo control eligieron a 200 mujeres a las que se había practicado una laparoscopia para una ligadura de trompas. En este grupo control sólo se diagnosticó endometriosis en 4 mujeres, lo que suponía una OR = 13. Se puede argumentar que la información sobre la endometriosis no es comparable en los casos y en los controles. No es lo mismo su búsqueda cuando la laparoscopia se realiza a causa de infertilidad que por ligadura de trompas. Probablemente, en el primer caso se buscó con más insistencia, dando lugar a un sesgo de información que sobrevaloraba la asociación. De todos modos, si este sesgo se corrige es muy probable que esta fuerte asociación disminuya, pero no que desaparezca. Por el contrario, si la asociación que se está estudiando es débil, la introducción de un sesgo puede alterar totalmente la interpretación de los resultados. Un estudio positivo puede pasar a ser negativo, y viceversa, dependiendo de la dirección del sesgo. Así pues, no sólo hay que valorar la posible existencia de errores sistemáticos, sino también su magnitud.

Un encuestador puede introducir un error de clasificación diferencial si conoce la hipótesis del estudio y la condición (de expuesto o enfermo, según el tipo de estudio) del entrevistado. En esta situación, puede obtener información de los individuos de forma diferente (más exacta o más exhaustiva) según a qué grupo de estudio pertenezcan.

**Ejemplo 29.10.** Aunque el objetivo de un estudio de casos y controles era la evaluación del efecto del consumo de alcohol sobre el infarto de miocardio, se explicaba a los participantes que se trataba de una investigación sobre factores de riesgo de la hospitalización. De esta forma, aunque los participantes sabían que habían sufrido un infarto de miocardio, no lo relacionaban con la hipótesis de estudio. Por otro lado, los investigadores obtenían información no sólo sobre el consumo de alcohol, sino también sobre otros factores, lo que servía, así mismo, para que los participantes no pudieran relacionar el consumo de alcohol como el único factor de riesgo en

estudio. De esta forma se minimizaba la probabilidad de un sesgo de memoria o de un sesgo debido al encuestador.

En los estudios de casos y controles, el sesgo debido al entrevistador se produce al recoger datos sobre la historia de exposición, dado que el conocimiento de si se está entrevistando a un caso o a un control puede condicionar la forma en que se recoge. Por la misma razón, en los estudios retrospectivos de cohortes también puede aparecer un sesgo debido al entrevistador cuando se recoge información sobre la exposición. En los estudios prospectivos de cohortes, el sesgo debido al entrevistador se puede dar en el momento de recoger información sobre la enfermedad (variable de respuesta). Este sesgo también puede aparecer en estudios experimentales en los que, por razones prácticas o éticas, no se utiliza un diseño controlado con placebo ni técnicas de enmascaramiento.

**Ejemplo 29.11.** En un estudio de casos y controles sobre el efecto protector de la aspirina frente al infarto agudo de miocardio se pueden introducir preguntas sobre la frecuencia de uso de otros medicamentos analgésicos de los que no se tenga constancia de que muestren ningún efecto sobre la enfermedad. Si los casos y los controles difieren en la frecuencia de uso de estos fármacos, se puede sospechar que se ha cometido un sesgo de información. Si son dos o más las personas que recogen la información, otra estrategia para evaluar la posible existencia de un sesgo de información es comparar los resultados obtenidos por cada una de ellas. Si se observan diferencias respecto a la frecuencia de exposición o en

otras variables, se pensará que existe un sesgo debido al entrevistador.

En el cuadro 29.2 se resumen los principales efectos de los errores de mala clasificación, tanto diferenciales como no diferenciales.

#### Cuadro 29.2 Principales efectos de la mala clasificación

- Cuando la enfermedad se clasifica como presente o ausente, la mala clasificación no diferencial hace que se diluya la asociación y en consecuencia se infraestima el efecto de interés
- Cuando la exposición se clasifica en dos categorías, la mala clasificación no diferencial infraestima la asociación
- Cuando la exposición se clasifica en más de dos categorías, la mala clasificación no diferencial puede sobreestimar o infraestimar la asociación
- La mala clasificación diferencial de la enfermedad o la exposición pueden conducir a un sesgo en cualquier dirección (sobreestimar o infraestimar la asociación)
- La mala clasificación de un factor de confusión conducirá a una ajuste incorrecto
- La mala clasificación de un factor de confusión puede dar lugar a estimaciones diferentes según las categorías de la variable de confusión, introduciendo una modificación espúrea del efecto

## BIBLIOGRAFÍA DE LOS EJEMPLOS

Meier P. The biggest public health experiment ever. En: Tanur JM, editor. Statistics: a guide to the unknown. 20.<sup>a</sup> ed. Holden-Day; 1978. p. 3-15.

Sartwell PE, Masi AT, Arthes FG, Greene GR, Smith HE. Tromboembolism and

oral contraceptives: an epidemiologic case-control study. Am J Epidemiol. 1969;90:365-80.

Strathy JH, Molgaard CA, Coulam CB, Melton 3rd LJ. Endometriosis and infertility: a laparoscopic study of

endometriosis among fertile and infertile women. Fertil Steril. 1982;38:667-72.

Werler M, Pober B, Nelson K, Holmes L. Reporting accuracy among mothers of malformed and nonmalformed infants. Am J Epidemiol.. 1989;129:415-21.

## BIBLIOGRAFÍA

- Berger VW, Exner DV. Detecting selection bias in randomized clinical trials. *Control Clin Trials.* 1999;20:319-27.
- Boutron I, Ravaud P. Classification systems to improve assessment of risk of bias. *J Clin Epidemiol.* 2012;65(3):236-8.
- Chalmers I. Comparing like with like: some historical mile-stones in the evolution of methods to create unbiased comparison groups in therapeutic experiments. *Int J Epidemiol.* 2001;30:1156-64.
- Chyou PH. Patterns of bias due to differential misclassification by case-control status in a case-control study. *Eur J Epidemiol.* 2007;22(1):7-17.
- Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol.* 2000;29:722-9.
- Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet.* 2002;359(19):248-52.
- Gurusamy KS, Gluud C, Nikolova D, Davidson BR. Assessment of risk of bias in randomized clinical trials in surgery. *Br J Surg.* 2009;96(4):342-9.
- Infante-Rivard C, Jacques L. Empirical study of parental recall bias. *Am J Epidemiol.* 2000;152:480-6.
- Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int J Epidemiol.* 2005;34(3):680-7.
- Lash TL, Silliman RA. A sensitivity analysis to separate bias due to confounding from bias due to predicting misclassification by a variable that does both. *Epidemiology.* 2000;11:544-9.
- Longford NT. Selection bias and treatment heterogeneity in clinical trials. *Stat Med.* 1999;18:1467-74.
- Lundberg M, Hallqvist J, Diderichsen F. Exposure-dependent misclassification of exposure in interaction analyses. *Epidemiology.* 1999;10:545-9.
- Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol.* 2002;31:422-9.
- Marcus SM. Assessing non-consent bias with parallel randomized and nonrandomized clinical trials. *J Clin Epidemiol.* 1997;50:823-8.
- Marshall RJ. An empirical investigation of exposure measurement bias and its components in case-control studies. *J Clin Epidemiol.* 1999;52:547-50.
- Nishioka SA, Gyorkos TW, Joseph L, Collet JP. Selection of subjects for hospital-based epidemiologic studies based on outward manifestations of disease. *Clin Invest Med.* 2001;24:299-303.
- Patten SB. Selection bias in studies of major depression using clinical subjects. *J Clin Epidemiol.* 2000;53:351-7.
- Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology.* 2003;14:459-66.
- Richiardi L, Boffetta P, Merletti F. Analysis of nonresponse bias in a population-based case-control study on lung cancer. *J Clin Epidemiol.* 2002;55:1033-40.
- Rockenbauer M, Olsen J, Czeizel AE, Pedersen L, Sorensen HT. EuroMAP Group. Recall bias in a case-control surveillance system on the use of medicine during pregnancy. *Epidemiology.* 2001;12:461-6.
- Selby JV, Friedman GD, Quesenberry CP, Weiss NS. Ascertainment bias in case-control studies of cancer screening. *J Clin Epidemiol.* 2001;54:215-6.
- Sturmer T, Thurigen D, Spiegelman D, Blettner M, Brenner H. The performance of methods for correcting measurement error in case-control studies. *Epidemiology.* 2002;13:507-16.
- Touloumi G, Pocock SJ, Babiker AG, Darbyshire JH. Impact of missing data due to selective dropouts in cohort studies and clinical trials. *Epidemiology.* 2002;13:347-55.
- White E. Design and interpretation of studies of differential exposure measurement error. *Am J Epidemiol.* 2003;157:380-7.

## Confusión y modificación del efecto

La finalidad del análisis de los resultados de un estudio es obtener la mejor estimación posible del efecto, la diferencia o la asociación de interés, para lo que se debe tener en cuenta la existencia de factores modificadores de dicho efecto y controlar la influencia de los factores de confusión.

### CARACTERÍSTICAS DEL FENÓMENO DE CONFUSIÓN

El fenómeno de confusión puede definirse como la distorsión que se produce en la estimación de la magnitud de la asociación entre el factor de estudio y la variable de respuesta, debida a la influencia de uno o varios factores externos.

Como ya se ha explicado en páginas anteriores, para producir confusión, una variable debe estar asociada tanto con el factor de estudio como con la variable de respuesta, y no debe formar parte de la cadena causal entre ellos. Por tanto, deben considerarse potenciales factores de confusión las variables por las que difieren los grupos que se comparan y que son pronósticas de la respuesta.

#### Debe estar asociado al factor de estudio y a la variable de respuesta

Si no existe dicha asociación, no se producirá un fenómeno de confusión. Retomando el ejemplo 2.12 sobre la relación entre el ejercicio físico y el infarto agudo de miocardio (IAM), los sujetos que realizan ejercicio y los que no, quizás diferirán en el consumo de agua, pero hasta el momento no se ha demostrado que la ingestión de agua esté relacionada con el riesgo de padecer un IAM. Así pues, el

consumo de agua no puede considerarse un factor de confusión de la asociación entre el ejercicio físico y la enfermedad.

#### No es necesario que la relación con la variable de respuesta sea causal

Un factor de confusión debe ser predictivo de la respuesta (es decir, estar asociado con ella), pero esta asociación no tiene que ser necesariamente causal. De hecho, la mayoría de los factores de confusión no son causa de la respuesta, sino simplemente marcadores que están correlacionados con el verdadero agente causal. En el ejemplo 2.13 sobre el riesgo de IAM en función del sexo, éste no es la verdadera causa, sino que simplemente expresa diferencias biológicas que podrían ser la causa de la enfermedad. La edad y el sexo son ejemplos claros de variables que están asociadas con multitud de enfermedades, pero que no son su causa directa y, a su vez, están relacionadas con muchos factores.

#### Debe ser un factor predictivo de la respuesta, independientemente del factor de estudio

Debe existir una asociación entre el factor de confusión y la respuesta también en los sujetos no expuestos al factor de estudio. En el ejemplo 2.12, si el ejercicio físico está inversamente relacionado con la probabilidad de desarrollar un IAM, el consumo de agua también lo estará, ya que los que realizan más ejercicio beben más agua. Sin embargo, como ya se ha comentado, el consumo de agua no se ha asociado con la enfermedad en los sujetos que no realizan ejercicio físico (no expuestos) y, por tanto,

no puede confundir la asociación entre el ejercicio físico y la enfermedad. Esta situación es muy distinta a la de la edad, el sexo o el consumo de tabaco, que son un factor de riesgo independiente entre los sujetos que no realizan ejercicio físico.

### No debe ser un paso intermedio en la cadena causal

El factor de confusión no puede ser un mero paso intermedio en la cadena causal. Esta distinción no es siempre clara y requiere conocimientos sobre los mecanismos causales de la respuesta.

Su identificación no debe basarse en la existencia de una diferencia estadísticamente significativa entre los grupos de estudio. Dado que la significación estadística depende tanto de la magnitud de la diferencia observada como del tamaño de la muestra, una diferencia muy pequeña e irrelevante puede ser estadísticamente significativa si el tamaño de los grupos es muy grande, y una diferencia de gran importancia puede ser estadísticamente no significativa si el número de sujetos incluidos es muy pequeño. Por lo tanto, la significación estadística no es un buen criterio para considerar que una variable puede ser un potencial factor de confusión. Lo que cuenta es la posible relevancia que pueda tener sobre los resultados la diferente distribución de una variable entre los grupos. Además, no debe olvidarse que múltiples diferencias de escasa importancia aparente pueden producir una distorsión del resultado clínicamente relevante, al actuar de forma conjunta.

**Ejemplo 30.1.** Supongamos un ensayo clínico aleatorio en el que 50 individuos son asignados al azar a dos grupos: uno que recibe el tratamiento A y otro que recibe el tratamiento B, y en el que se observa un desequilibrio en el porcentaje de mujeres que han quedado asignadas a cada grupo (el 60% en el tratamiento A frente al 40% en el tratamiento B). Esta diferencia no es estadísticamente significativa ( $p=0,20$ ), debido, probablemente, al reducido tamaño de la muestra, pero si el sexo fuera una variable pronóstica del resultado, la magnitud de la diferencia podría ser lo suficientemente relevante como para distorsionar el resultado del estudio.

Además de identificar los potenciales factores de confusión, es conveniente tener en cuenta su dirección, es decir, cómo pueden afectar a los resultados.

**Ejemplo 30.2.** Supongamos un estudio que evalúa las diferencias en las concentraciones de colesterol ligado a las lipoproteínas de alta densidad (cHDL) entre enfermos con sarcoidosis activa y los que no presentan actividad. Se observa que los pacientes con enfermedad

activa tienen unas concentraciones de cHDL más bajas y que, a su vez, el porcentaje de mujeres es más alto que en el grupo con enfermedad inactiva. Es conocido, por estudios previos, que las mujeres tienen unas concentraciones de cHDL superiores a las de los hombres. Si no se ajustaran las diferencias observadas de cHDL en función del sexo, se obtendría como resultado un efecto observado menor que el verdadero. Por el contrario, en otras ocasiones, no ajustar por factores de confusión puede suponer obtener en el estudio una asociación mayor que la verdadera.

Desde un punto de vista práctico, la presencia de un fenómeno de confusión puede detectarse comprobando si el análisis bruto (sin tener en cuenta el factor de confusión) y el ajustado (controlando el potencial fenómeno de confusión) conducen a resultados que difieren de forma clínicamente relevante. La evaluación de la confusión no se basa, pues, en una prueba estadística, sino en la valoración de lo relevante que pueda considerarse la diferencia entre ambas estimaciones. Aunque es difícil fijar un criterio único para concluir que ha existido confusión, ya que depende de cada situación concreta, algunos autores lo establecen en un cambio de un 15-20% en la estimación del efecto de interés. Sin embargo, debe tenerse en cuenta que lo importante no es tanto identificar si ha existido confusión, sino, en caso de que se haya producido, obtener una estimación no distorsionada de la magnitud del efecto del factor de estudio sobre la variable de respuesta, es decir, que se haya controlado su influencia (estimación ajustada).

Aunque en la fase de análisis puede ajustarse por las variables de confusión, no hay que olvidar que ya en el diseño se debe tener en cuenta qué variables pueden actuar como tales, para medirlas. Esta identificación depende del conocimiento que se tenga de los mecanismos causales, así como de estudios previos que se hayan realizado sobre una hipótesis similar. La información que se recoge sobre ellos debe ser lo suficientemente detallada como para permitir su adecuado control en el análisis.

**Ejemplo 30.3.** Supongamos que un posible factor de confusión de la asociación entre el ejercicio físico y el riesgo de desarrollar un IAM es el consumo de tabaco. En un estudio sobre esta asociación no sería suficiente recoger información sobre si un individuo fuma, ya que no podría controlarse toda la distorsión debida a la cantidad de tabaco consumido, de manera que quedaría un fenómeno de confusión residual sin controlar, al considerar que todos los fumadores tienen el mismo riesgo, sin tener en cuenta el gradiente de riesgo asociado a la cantidad de tabaco consumido.

Dado que los factores de confusión son una amenaza constante a la validez de los datos de un estudio, no sólo es necesario considerar qué variables pueden confundir una asociación, sino también medirlas de forma válida y precisa.

## AJUSTE POR FACTORES DE CONFUSIÓN

Aunque el ajuste puede realizarse en la fase de análisis, también pueden utilizarse estrategias en el diseño del estudio destinadas a evitar o reducir la confusión (cuadro 30.1). Estos procedimientos pueden utilizarse en diferentes combinaciones para aumentar el grado de control.

### Técnicas de control de factores de confusión en el diseño

Dado que para que una variable actúe como un factor de confusión es necesario que se distribuya de forma desigual entre los grupos de estudio, las técnicas que buscan conseguir su distribución equilibrada tienden a prevenir su aparición. Sin embargo, no todas estas técnicas pueden utilizarse en todos los tipos de estudio.

### Asignación aleatoria

En los ensayos clínicos, la asignación aleatoria de los sujetos a los grupos de estudio tiende a conseguir grupos similares para todas las variables pronósticas del resultado, ya que el único factor que interviewe en su distribución es el azar. De todos modos, aunque tiende a conseguir grupos similares, no lo asegura. Cuando se trata de grupos pequeños, la probabilidad de que las variables se distribuyan desigualmente y puedan actuar como factores de

Cuadro 30.1 Métodos para controlar factores de confusión

En la fase de diseño:

- Asignación aleatoria
- Restricción

En la fase de diseño y análisis:

- Emparejamiento

En la fase de análisis:

- Análisis estratificado
- Análisis multivariante

confusión es mayor (ejemplo 30.1). Por ello, debe comprobarse la comparabilidad de los grupos y, en caso de que existan potenciales factores de confusión, ajustar el análisis por ellos.

### Restricción

Una forma de conseguir grupos comparables respecto a la distribución de una determinada variable es restringiendo la selección de los sujetos del estudio. Por ejemplo, si el sexo es un potencial factor de confusión, se puede decidir incluir sólo a mujeres. En ocasiones, la restricción reduce la influencia de una variable, pero no la elimina totalmente. Por ejemplo, aunque en un estudio se restrinja la inclusión a los sujetos de edades entre 40 y 60 años, si existe un gradiente de riesgo asociado a la edad (el riesgo de los sujetos de 60 años es diferente del de los de 40 años) y esta variable se distribuye de forma desequilibrada en los grupos de estudio, no se ha eliminado completamente el efecto de la edad, de modo que puede quedar cierto fenómeno de confusión residual.

El uso de criterios de selección restrictivos tiene el inconveniente de que reduce el número de sujetos candidatos, lo que puede suponer una dificultad para alcanzar el tamaño de la muestra necesario para el estudio en un tiempo razonable.

Además, limita la capacidad de generalización de los resultados, ya que se han utilizado criterios de selección restrictivos, y no permite evaluar el efecto de las variables no consideradas (p. ej., si el estudio sólo incluye mujeres, no podrá analizarse si el efecto del factor de estudio es diferente en los varones).

### Emparejamiento

Esta técnica implica que los sujetos del grupo de estudio y los del de control se emparejan en función de los valores de un determinado factor de confusión. Por ejemplo, si en un estudio de casos y controles sobre el ejercicio físico y el riesgo de IAM, el consumo de tabaco es un potencial factor de confusión, se emparejará cada caso con un control, o más de uno, que tenga el mismo hábito tabáquico.

El emparejamiento fuerza que la distribución de los factores de confusión sea idéntica entre los distintos grupos del estudio, lo que obliga a utilizar en el análisis técnicas estadísticas que tengan en cuenta la naturaleza apareada de los datos para poder controlar el efecto del factor de confusión.

Esta técnica tiene limitaciones. Puede ser muy difícil encontrar los controles adecuados, especialmente si se desea emparejar por muchas variables. Además, imposibilita evaluar el efecto de las variables por las que se ha emparejado.

El desarrollo de las técnicas de análisis de tipo multivariante, que permiten controlar los factores de confusión sin necesidad de recurrir al emparejamiento, ha disminuido el atractivo de este método.

## Técnicas de ajuste por factores de confusión

El procedimiento de control de factores de confusión en la fase de análisis se conoce como ajuste estadístico. La estrategia de análisis debe prever y especificar las técnicas que se utilizarán con esta finalidad.

**Ejemplo 30.4.** Müllner et al (2002) revisaron 537 artículos originales publicados en 34 revistas médicas, de los que 169 (32%) informaban que habían realizado algún tipo de ajuste por factores de confusión. En un 45% de los casos no quedaba claro cómo se habían controlado las variables continuas o de múltiples categorías, es decir, si se habían clasificado en intervalos o se habían agregado categorías, de manera que no podía evaluarse si se había controlado todo el fenómeno de confusión que podían haber producido. La conclusión que se extrae es que los artículos frecuentemente no informan de los métodos de ajuste o lo hacen inadecuadamente.

## Análisis estratificado

La estratificación permite evaluar la asociación entre el factor de estudio y la variable de respuesta dentro de categorías homogéneas (estratos) de una tercera variable. Si se evalúa la asociación entre ejercicio físico e IAM en los hombres y en las mujeres por separado, su estimación no estará interferida por

la variable sexo, ya que en cada estrato no existirá variabilidad por este factor, de forma que se habrá eliminado su potencial efecto de confusión. Si no existen diferencias en las estimaciones obtenidas en cada estrato, es preferible presentar los resultados como una estimación conjunta del efecto entre el ejercicio físico y el IAM que haya tenido en cuenta la influencia del sexo, es decir, ajustada (ponderada según los estratos).

**Ejemplo 30.5.** Supongamos un estudio sobre la asociación entre ejercicio físico e IAM, en el que se considera que el consumo de tabaco puede ser un potencial factor de confusión. Para controlar su efecto, se divide esta variable en categorías homogéneas (estratos), en cada una de las cuales se analiza el efecto del ejercicio físico sobre la enfermedad. Aunque las categorías de consumo de tabaco podrían ser múltiples (p. ej., nunca fumador, fumador de menos de 10 cigarrillos/día, fumador de más de 10 cigarrillos/día y ex fumador), para ilustrar este ejemplo se han considerado tan sólo dos estratos: no fumador y fumador.

De los datos de la tabla 30.1 se desprende que el ejercicio físico es un factor protector del IAM tanto en fumadores como en no fumadores. La *odds ratio* (OR) calculada en cada uno de los estratos es una estimación de la asociación entre el ejercicio físico y el IAM en fumadores y en no fumadores. Dado que ambos resultados son muy similares, es preferible obtener una estimación global ponderada que tenga en cuenta la información de todos los individuos incluidos en el estudio. Cuando se trata de tablas  $2 \times 2$ , el procedimiento más utilizado para realizar

**Tabla 30.1 Ejemplo de análisis estratificado. Estudio hipotético de la asociación entre ejercicio físico (EF) e infarto agudo de miocardio (IAM), con estratificación por el consumo de tabaco (ejemplo 30.5)**

		Casos	Controles	
<b>Análisis de datos crudos</b>				
	EF sí	222	121	
	EF no	194	69	OR=0,65
<b>Análisis estratificado</b>				
Fumadores	EF sí	112	79	
	EF no	27	16	OR=0,84
No fumadores	EF sí	110	42	
	EF no	167	53	OR=0,83
OR ajustada (Mantel-Haenszel)=0,834.				

esta ponderación es el de Mantel y Haenszel. En este estudio, la estimación del riesgo de IAM asociado a la práctica de ejercicio físico, ajustada por el consumo de tabaco, es de 0,834. La diferencia entre este valor y el obtenido con los datos crudos ( $OR=0,65$ ) permite cuantificar el fenómeno de confusión por el consumo de tabaco.

Como ya se ha comentado, la magnitud de un efecto de confusión se evalúa observando el grado de discrepancia entre las estimaciones crudas y las ajustadas, y valorando no su significación estadística, sino su relevancia clínica. En el ejemplo anterior, como las estimaciones cruda y ajustada difieren (0,65 y 0,83, respectivamente), si se considera que la diferencia es clínicamente relevante, puede decirse que ha existido efecto de confusión debido a la variable consumo de tabaco. La mejor estimación es la ajustada, ya que ha controlado este fenómeno de confusión.

El análisis estratificado es una técnica sencilla, asequible para los investigadores y los lectores de revistas biomédicas, que permite comprender fácilmente las interrelaciones entre el factor de estudio, la respuesta y las variables que pueden actuar como factores de confusión y/o modificadoras del efecto. Su principal limitación es la dificultad para estratificar por múltiples variables. Por ejemplo, si se desea controlar por tres variables (de 2, 3 y 5 categorías, respectivamente), existirán 30 posibles estratos y, aunque en el estudio se haya incluido un número importante de sujetos, es muy posible que en algunos estratos no exista el número de efectivos suficiente para estimar un efecto con una precisión aceptable.

## Análisis multivariante

El análisis multivariante supera esta limitación del análisis estratificado y permite estimar de forma eficiente el efecto de un factor de estudio sobre una enfermedad, ajustado por varios potenciales factores de confusión, simultáneamente.

Su uso implica la construcción de un modelo matemático, en el que deben identificarse la variable dependiente (variable de respuesta), la variable independiente principal (factor de estudio) y el resto de las variables independientes (factores de confusión que controlar), y se determinan los valores de los coeficientes de regresión de cada una de las variables independientes. El coeficiente del factor de estudio corresponde a la estimación de su efecto sobre la variable de respuesta, ajustado por el resto de las variables independientes incluidas en el modelo.

**Ejemplo 30.6.** Supongamos un ensayo clínico hipotético en el que se compara la eficacia de un tratamiento

antihipertensivo con un placebo. Se utiliza como variable de respuesta la disminución de las cifras de presión arterial diastólica (PAD). Los resultados se resumen en la tabla 30.2. En el grupo con tratamiento activo, se ha observado un descenso de 25 mmHg, mientras que ha sido sólo de 10 mmHg en el grupo placebo, por lo que el efecto del tratamiento puede cuantificarse en una disminución de 15 mmHg (en la ecuación de regresión obtenida con los datos crudos, el coeficiente de la variable tratamiento es 15).

Dado que los investigadores consideran que el sexo podría ser un factor de confusión, controlan su efecto, obteniendo una estimación del efecto del tratamiento de 5 mmHg tanto en hombres como en mujeres. Si se hubiera realizado un análisis multivariante, se habría obtenido una ecuación de regresión en la que el coeficiente de la variable tratamiento sería 5, una vez controlado el efecto de la variable sexo (también incluida en la ecuación). El coeficiente de la variable sexo indicaría la diferencia en la respuesta observada entre hombres y mujeres (obsérvese que la disminución de las cifras de PAD en los hombres ha sido 20 mmHg superior a la conseguida en las mujeres).

En este ejemplo puede decirse que ha existido un fenómeno de confusión debido al sexo, ya que los coeficientes de regresión de la variable tratamiento, calculados en el modelo con datos crudos y en el modelo que incluye también la variable sexo, difieren (15 y 5, respectivamente). Dado que el objetivo del estudio es estimar el efecto del tratamiento activo, el mejor resultado es el ajustado por el sexo, es decir, 5 mmHg.

Existen diferentes técnicas de análisis multivariante. La elección de la más adecuada depende del diseño del estudio, de la naturaleza de las variables y de las interrelaciones entre el factor de estudio, la variable de respuesta y los restantes factores incluidos en el modelo, así como de las condiciones específicas de aplicación de cada una de ellas (anexo 8).

## Problemas del ajuste por factores de confusión

La mayoría de los estudios clínicos y epidemiológicos presentan estimaciones del efecto ajustadas por las variables consideradas como potenciales factores de confusión. Sin embargo, el hecho de haber ajustado por algunas variables no garantiza que se haya eliminado cualquier efecto de confusión. Pueden existir algunos problemas relacionados con la información de las variables de confusión o con el uso de técnicas estadísticas inadecuadas, que pueden conducir a fallos en el ajuste (cuadro 30.2).

**Tabla 30.2 Ejemplo de análisis multivariante. Ensayo clínico hipotético que compara un fármaco antihipertensivo con placebo. La variable de respuesta es la disminución de las cifras de presión arterial diastólica (dif. PAD). Se estima el efecto del tratamiento con ajuste por la variable sexo (ejemplo 30.6)**

	Dif. PAD	Efecto del tratamiento
<b>Análisis de datos crudos</b>		
Tratamiento activo	25 mmHg	
Placebo	10 mmHg	15 mmHg
Ecuación de regresión lineal Dif. PAD = 10 + 15 · tratamiento Códigos: tratamiento (0 = placebo; 1 = tratamiento activo)		
<b>Análisis estratificado</b>		
Hombres	Tratamiento activo	30 mmHg
	Placebo	25 mmHg
Mujeres	Tratamiento activo	10 mmHg
	Placebo	5 mmHg
Ecuación de regresión lineal Dif. PAD = 5 + 5 · tratamiento + 20 · sexo Códigos: tratamiento (0 = placebo; 1 = tratamiento activo) sexo (0 = mujer; 1 = hombre)		

## Información inadecuada sobre la variable de confusión

### Agrupación inadecuada de los datos

Cuando las categorías que se crean son demasiado amplias y dentro de cada una de ellas existe cierta heterogeneidad o gradiente de riesgo, el ajuste no controla todo el fenómeno de confusión y queda cierto efecto residual (ejemplo 30.3).

### Utilización de medidas indirectas poco válidas

Dadas las dificultades para medir algunos de los fenómenos de interés, suelen utilizarse medidas indirectas. Cuando se ajusta por ellas, no se controla todo el efecto de dicho fenómeno, y queda cierta confusión residual.

**Ejemplo 30.7.** Supongamos que se quiere estudiar la relación entre la dieta y el cáncer de mama y se desea ajustar la estimación del efecto por la clase socioeconómica. Si se utiliza como medida indirecta de esta variable el grado de estudios, dado que éste no la representa adecuadamente, el ajuste de los datos sólo será parcial.

## Cuadro 30.2 Problemas en el ajuste por factores de confusión

- Información inadecuada sobre la variable de confusión
  - Agrupación inadecuada de los datos
  - Utilización de medidas indirectas poco válidas
  - Mala clasificación de las variables de confusión
- Variables muy correlacionadas
- Existencia de modificación de efecto
- Variable intermedia en la cadena causal
- Variables que no son factores de confusión
- Pruebas estadísticas inadecuadas
- Selección automática de variables en un modelo de regresión

### Mala clasificación de las variables de confusión

Cuando las variables de confusión se miden de forma imprecisa o poco válida, sus relaciones con el resto de las variables no se estiman correctamente, de forma que, al ajustar por ellas, no se controla totalmente el fenómeno de confusión, y también puede quedar un efecto de confusión residual.

**Ejemplo 30.8.** Supongamos un estudio que desea estimar el efecto del consumo de tabaco sobre el riesgo de padecer un IAM. Los investigadores sospechan

que la asociación puede ser total o parcialmente explicada por la influencia de la práctica de ejercicio físico. Asumamos que el 10% de los hombres no realiza actividad física. La tabla 30.3a muestra los resultados que se obtendrían cuando no hay mala clasificación.

Para ajustar correctamente por un factor de confusión, los instrumentos de medida deberían ser perfectos. En la práctica, la medición del grado de actividad física es difícil y puede cambiar en el tiempo. Además, pueden existir errores de información debidos al instrumento de medida utilizado, al propio observador o, sencillamente, a que los individuos falseen las respuestas sobre sus hábitos.

Supongamos que el 50% de los hombres que practican algún deporte son clasificados como que no lo practican, mientras que el 5% de los que no hacen ejercicio físico son clasificados como que sí. En la tabla 30.3b se presentan los resultados que se obtendrían con una mala clasificación no diferencial de estas características. En esta situación, en ambos estratos se sobreestima el efecto del consumo de tabaco sobre el riesgo de IAM (que era de OR= 1,5). Así pues, aunque aparentemente se esté ajustando por un factor de confusión, en realidad se está introduciendo un sesgo. Además, la mala clasificación no diferencial del factor de confusión ha conducido

a una estimación del efecto diferente en cada uno de los estratos, de forma que podría sugerir erróneamente que la actividad física modifica el efecto del consumo de tabaco sobre el IAM. La situación llega a ser mucho más compleja cuando hay más de un factor de confusión que no se ha medido correctamente.

### Variables muy correlacionadas

Aunque para que una variable actúe como un factor de confusión debe estar asociada con el factor de estudio, si ambas variables están muy relacionadas, puede ser difícil distinguir sus efectos de forma precisa.

Intuitivamente, el ajuste implica estimar el efecto del factor de estudio sobre la variable de respuesta, mientras se mantiene fijo el valor del factor de confusión. Por lo tanto, si el factor de estudio y el de confusión están muy relacionados, prácticamente todos los sujetos con un valor determinado de la variable de confusión tendrán el mismo valor del factor de estudio y, en consecuencia, será difícil estimar su efecto sobre la variable de respuesta.

Si se ajusta por esta variable, el resultado será una estimación del efecto imprecisa, de tal forma que los intervalos de confianza de la estimación no ajustada serán más estrechos que los de la ajustada.

**Tabla 30.3 Efecto de los errores de medida del factor de confusión sobre el ajuste. Estudio hipotético del efecto del consumo de tabaco sobre el riesgo de infarto agudo de miocardio (IAM) ajustado por la realización de ejercicio físico (EF) (ejemplo 30.8)**

		IAM	no IAM	Total	OR
a) Situación en que no existe mala clasificación					
EF sí	Fuma	120	680	800	
	No fuma	20	180	200	1,5
EF no	Fuma	54	1.746	1.800	
	No fuma	144	7.056	7.200	1,5
Total		338	9.662	10.000	
b) Situación en que existe mala clasificación (el 50% de los hombres que realizan ejercicio físico son clasificados como que no lo realizan, mientras que el 5% de los que no lo hacen son clasificados como que sí)					
EF sí	Fuma	63*	427	490	
	No fuma	17	443	460	3,8
EF no	Fuma	111	1.999	2.110	
	No fuma	147	6.793	6.490	2,7
Total		338	9.662	10.000	

\* $63 = 120 - (0,5 \times 120) + (0,05 \times 54)$ .

El resto de las casillas se calculan de la misma forma aplicando los porcentajes correspondientes.

## Existencia de una modificación del efecto

La asunción que subyace cuando se utiliza la estimación ajustada es que el factor de estudio tiene el mismo efecto en cada uno de los estratos o categorías de la variable de confusión. Cuando existe una modificación del efecto, y los resultados son diferentes en cada estrato, la presentación de una única estimación ajustada no es adecuada.

## Variable intermedia en la cadena causal

No es apropiado ajustar por una variable que es un paso intermedio en la cadena causal entre el factor de estudio y la variable de respuesta. Si se hiciera dicho ajuste, se obtendría una infraestimación de la verdadera asociación, ya que se eliminaría de la estimación del efecto la parte mediada a través de dicha variable intermedia.

**Ejemplo 30.9.** Supongamos que en un estudio que trata de estimar el efecto del consumo de tabaco sobre la mortalidad infantil, se ajuste por el peso del recién nacido. El bajo peso en los recién nacidos es una de las consecuencias del consumo de tabaco y, a su vez, puede ser una causa de mortalidad infantil. Al ajustar por el bajo peso, parte del efecto del tabaco (exposición) quedaría neutralizado por dicho ajuste (se eliminaría la parte del efecto del tabaco que actúa a través del bajo peso), de forma que se infraestimaría su verdadera magnitud.

Una de las condiciones que debe reunir una variable para ser considerada un potencial factor de confusión es no ser un paso intermedio en la cadena causal. Con frecuencia, es difícil distinguir si una variable lo es, por lo que se requiere un conocimiento profundo de los mecanismos biológicos por los que una exposición puede causar una enfermedad.

**Ejemplo 30.10.** Si se evalúa el efecto del consumo moderado de alcohol sobre el riesgo de padecer un IAM, una variable que puede considerarse un potencial factor de confusión es la concentración de cHDL. En algunos estudios se ha observado que el consumo moderado de alcohol aumenta las concentraciones plasmáticas de cHDL y que, a mayor concentración de cHDL, menor riesgo de padecer un IAM. Se ha postulado que uno de los mecanismos por los que el consumo moderado de alcohol reduce el riesgo de IAM es precisamente el aumento de cHDL. Si se asume esta posible cadena causal, las concentraciones de cHDL no deben considerarse un factor de confusión, y el efecto del consumo de alcohol sobre el IAM no debe ajustarse por ellas. Por otro lado, si

es de interés evaluar hasta qué punto el consumo de alcohol tiene un efecto sobre el IAM por mecanismos distintos al aumento del cHDL, entonces el análisis debería contemplar el cHDL como una variable que controlar.

## Variables que no son factores de confusión

El ajuste por variables que no son verdaderos factores de confusión no sesga la estimación, pero la hace más imprecisa (fenómeno que se conoce como sobreajuste).

El uso generalizado de programas informáticos que permiten realizar análisis estadísticos complejos con multitud de variables, de forma rápida y cómoda, facilita que algunos investigadores incluyan en el análisis gran cantidad de variables, algunas de las cuales no son relevantes para la relación que se estudia. Aunque las técnicas de análisis multivariante tienen ventajas respecto a las del análisis estratificado, para que sean realmente útiles se deben tener en cuenta dos condiciones básicas: el objetivo del estudio debe estar bien definido, especificando cuál es la variable de estudio y cuál la de respuesta, cuáles los potenciales factores de confusión y cuáles las potenciales variables modificadoras del efecto; y hay que incluir en el modelo sólo aquellas variables que se conoce o se sospecha firmemente que pueden llevar a confusión o modificar el efecto.

## Pruebas estadísticas inadecuadas

Si no se utilizan las pruebas estadísticas correctas, no se controla adecuadamente el fenómeno de confusión. Por ejemplo, si en un estudio se utiliza el emparejamiento, pero en el análisis no se tiene en cuenta la naturaleza apareada de los datos, la estimación del efecto resultará sesgada.

## Selección de variables en un modelo de regresión

La finalidad del análisis estadístico es obtener la mejor estimación (no sesgada y lo más precisa posible) del efecto del factor de estudio. En esta situación, el uso de un algoritmo automático para la selección de las variables del modelo multivariante puede no incluir variables importantes porque individualmente no son estadísticamente significativas; sin embargo, estas variables tomadas en conjunto podrían ser responsables de una parte sustancial del efecto estudiado, y por tanto serían variables relevantes que controlar.

## EVALUACIÓN DE LA MODIFICACIÓN DEL EFECTO

En las situaciones anteriores, los efectos observados en los diferentes estratos eran similares, por lo que tenía sentido obtener una única estimación ajustada. Sin embargo, cuando no lo son y puede considerarse que existe una modificación del efecto, no es adecuado calcular un único resultado global, sino que lo importante es precisamente describir que el efecto es diferente según las diversas categorías de otras variables. A diferencia del fenómeno de confusión, la modificación del efecto se evalúa estadísticamente, y puede decirse que existe cuando las estimaciones de cada uno de los estratos presentan una diferencia estadísticamente significativa. Las técnicas que se utilizan para evaluar su existencia son también el análisis estratificado y el multivariante.

### Análisis estratificado

El análisis estratificado se realiza de la misma forma que en el caso del ajuste por factores de confusión: calculando la estimación del efecto en cada uno de los estratos definidos por las categorías de la variable cuya influencia se desea evaluar.

Pero, a diferencia de lo que ocurre con el fenómeno de confusión, la evaluación de la existencia de modificación del efecto en un análisis estratificado se realiza mediante pruebas estadísticas específicas, como la de Welch, ya que se trata de determinar si las diferencias entre las estimaciones obtenidas en cada estrato

pueden ser causadas por la variabilidad aleatoria o si, por el contrario, reflejan diferencias reales.

**Ejemplo 30.11.** Supongamos un estudio de casos y controles (tabla 30.4) que evalúa la asociación entre la práctica de ejercicio físico y el riesgo de sufrir un IAM, y que se desea evaluar si esta asociación es diferente según el sexo de los sujetos.

Los datos crudos sugieren que el ejercicio físico disminuye el riesgo de IAM ( $OR = 0,64$ ). Sin embargo, al estratificar, en los hombres se observa un efecto protector ( $OR = 0,53$ ), mientras que en las mujeres no ( $OR = 1,19$ ). Si se realizara una prueba estadística para comparar ambas estimaciones, se obtendría un resultado estadísticamente significativo, que indicaría que el sexo es una variable modificadora del efecto, y que la asociación entre el ejercicio físico y el IAM es diferente en los hombres y en las mujeres.

Si se hubiera realizado una estimación ajustada, se hubiera obtenido un valor de  $OR = 0,64$ , similar al calculado con los datos crudos, que parecería indicar que existe un efecto protector en todos los sujetos, sin poner de manifiesto que este efecto solamente se ha observado en los hombres. Por lo tanto, cuando existe una modificación del efecto, es preferible presentar los resultados por estratos y no utilizar una medida global ajustada, que podría conducir a interpretaciones erróneas.

### Análisis multivariante

Cuando se utiliza una técnica de análisis multivariante para evaluar la presencia de modificación del efecto, debe incluirse en el modelo no sólo la

**Tabla 30.4** Modificación de efecto en un análisis estratificado. Estudio hipotético de la asociación entre ejercicio físico (EF) e infarto agudo de miocardio (IAM), estratificado por sexo (ejemplo 30.11)

		Casos	Controles	
<b>Análisis de datos crudos</b>				
	EF sí	190	266	$OR = 0,64$
	EF no	176	157	
<b>Análisis estratificado</b>				
Hombres	EF sí	141	208	$OR = 0,53$
	EF no	144	112	IC 95% = 0,38 a 0,73
Mujeres	EF sí	110	42	$OR = 1,19$
	EF no	167	53	IC 95% = 0,65 a 2,16
OR ajustada (Mantel-Haenszel) = 0,64.				

variable que potencialmente puede producir dicho efecto, sino también un término que refleje la interacción entre esta variable y el factor de estudio. Dado que muchos programas estadísticos no evalúan la existencia de modificación del efecto de forma automática, es necesario crear intencionadamente los términos de interacción antes del análisis.

Puede decirse que existe modificación del efecto cuando el coeficiente de regresión del término de interacción es estadísticamente significativo (diferente de 0). En este caso, la estimación del efecto no corresponde solamente al valor del coeficiente de regresión del factor de estudio, sino que también incluye el valor de los coeficientes de los términos de interacción.

**Ejemplo 30.12.** Recuérdese el ejemplo 30.6 que presentaba un ensayo clínico hipotético en el que se comparaba la eficacia de un tratamiento antihipertensivo con un placebo, utilizando como variable de respuesta la disminución de las cifras de PAD.

Supongamos que los investigadores desean estudiar la influencia de la presencia de diabetes mellitus sobre la eficacia del tratamiento. Al estratificar (tabla 30.5), se observa que el efecto del tratamiento en relación con el placebo es de 20 mmHg en los

diabéticos y de sólo 5 mmHg en los no diabéticos. Dado que la diferencia entre ambas estimaciones es estadísticamente significativa, se concluye que la diabetes modifica el efecto del tratamiento, es decir, que la eficacia es diferente en diabéticos y en no diabéticos.

Si se hubiera realizado un análisis multivariante, además de los términos *tratamiento* y *diabetes*, se habría incluido en el modelo un término de interacción *tratamiento*  $\times$  *diabetes*. En la ecuación obtenida, este término tiene un coeficiente de 15, que es estadísticamente significativo, e indica la existencia de interacción entre el tratamiento y la presencia de diabetes. Si bien el coeficiente de la variable tratamiento es 5, el efecto del tratamiento difiere según la presencia de diabetes, por lo que su estimación es  $5 \times (15 \times \text{diabetes})$ . Así, en los sujetos no diabéticos (*diabetes* = 0), el efecto es  $5 \times 0 = 5$  mmHg, mientras que en los diabéticos (*diabetes* = 1) es  $5 \times 15 = 20$  mmHg. De esta forma, el análisis multivariante permite expresar, en una sola ecuación, los resultados de los diferentes estratos.

Los fenómenos de confusión y de modificación del efecto son muy distintos tanto en la información que proporcionan como en la forma de interpretarla. Si una variable actúa como un factor de confusión depende de cómo se distribuye entre los grupos de

**Tabla 30.5** Modificación de efecto en un análisis multivariante. Ensayo clínico hipotético que compara un fármaco antihipertensivo con placebo. La variable de respuesta es la disminución de las cifras de presión arterial diastólica (dif. PAD). Se evalúa la existencia de modificación de efecto debida a la diabetes (ejemplo 30.12)

	Dif. PAD	Efecto del tratamiento
<b>Análisis de datos crudos</b>		
Tratamiento activo	25 mmHg	
Placebo	10 mmHg	15 mmHg
Ecuación de regresión lineal Dif. PAD = 10 + 15 · tratamiento Códigos: tratamiento (0 = placebo; 1 = tratamiento activo)		
<b>Análisis estratificado</b>		
Diabéticos	Tratamiento activo	35 mmHg
	Placebo	15 mmHg
No diabéticos	Tratamiento activo	10 mmHg
	Placebo	5 mmHg
Ecuación de regresión lineal Dif. PAD = 5 + 5 x tratamiento + 10 x diabetes + 15 x tratamiento $\times$ diabetes Códigos: tratamiento (0 = placebo; 1 = tratamiento activo); diabetes (0 = no; 1 = sí)		

estudio y de si el efecto observado del factor de estudio sobre la variable de respuesta puede explicarse total o parcialmente por esta distribución desigual. Por lo tanto, una variable puede actuar como factor de confusión en un estudio experimental, pero no en otro, dependiendo de cómo quede distribuida tras el proceso de asignación. En cambio, en los estudios observacionales, al estar asociadas a la exposición y a la variable de respuesta, los factores de confusión siempre actúan como tales cuando no se ha utilizado una técnica para prevenir su efecto (restricción o

emparejamiento). Cuando una variable se comporta como un factor de confusión, el interés reside en controlar su efecto.

En cambio, la modificación del efecto refleja una característica de la relación entre el factor de estudio y la variable de respuesta, ya que la magnitud de esta relación depende de una tercera variable. Por tanto, la modificación del efecto debe ser descrita, pero no controlada. Además, puede proporcionar claves interesantes (hipótesis de trabajo) sobre los mecanismos causales de dicha relación.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

Müllner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med.* 2002;136:122-6.

## BIBLIOGRAFÍA

- Cobo E, Buekens P. Necesidad y limitaciones del ajuste. *Med Clin (Barc).* 1990;95:702-8.
- Cole SR, Hernan MA. Fallibility in estimating direct effects. *Int J Epidemiol.* 2002;31:163-5.
- De Graaf MA, Jager KJ, Zoccali C, Dekker FW. Matching, an appealing method to avoid confounding? *Nephron Clin Pract.* 2011;118(4):c315-8.
- Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health.* 2001;22:189-212.
- Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet.* 2002;359:248-52.
- Groenwold RH, Klungel OH, Grobbee DE, Hoes AW. Selection of confounding variables should not be based on observed associations with exposure. *Eur J Epidemiol.* 2011;26(8):589-93.
- Groenwold RH, Nelson DB, Nichol KL, Hoes AW, Hak E. Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research. *Int J Epidemiol.* 2010;39(1):107-17.
- Hallqvist J, Ahlbom A, Diderichsen F, Reuterwall C. How to evaluate interaction between causes: a review of practices in cardiovascular epidemiology. *J Intern Med.* 1996;239:377-82.
- Hernan MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol.* 2002;155:176-84.
- James H, Dominici F, Zeger S. On quantifying the magnitude of confounding. *Biostatistics.* 2010;11(3):572-82.
- Julious SA, Mullee MA. Confounding and Simpson's paradox. *BMJ.* 1994;309:1480-1.
- Kleinbaum DG. Logistic regression: a self-learning text. Nueva York: Springer-Verlag; 1994.
- Kleinbaum DG. Survival analysis: a self-learning text. Nueva York: Springer-Verlag; 1996.
- Leon DA. Failed or misleading adjustment for confounding. *Lancet.* 1993;342:479-81.
- MacLure M, Schneeweiss S. Causation of bias: the episcope. *Epidemiology.* 2001;12:114-22.
- Shapiro S. Bias in the evaluation of low-magnitude associations: an empirical perspective. *Am J Epidemiol.* 2000;151:939-45.
- Shahar E, Shahar DJ. On the causal structure of information bias and confounding bias in randomized trials. *J Eval Clin Pract.* 2009;15(6):1214-6.
- Silva Ayçaguer LC. Excursión a la regresión logística en ciencias de la salud. Madrid: Díaz de Santos; 1995.
- Silverman SL. From randomized controlled trials to observational studies. *Am J Med.* 2009;122(2):114-20.
- Smith GD, Phillips AN. Confounding in epidemiological studies: why «independent» effects may not be all they seem. *BMJ.* 1992;305:757-9.

- Sonis J. A closer look at confounding. *Fam Med.* 1998;30:584-8.
- Sonis J, Hertz-Pannier I. Assessing the presence of confounding. *Fam Med.* 1996;28:462-3.
- Sørensen HT, Lash TL, Rothman KJ. Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. *Hepatology.* 2006;44(5):1075-82.
- Tripepi G, Jager KJ, Dekker FW, Zoccali C. Stratification for confounding-part 1: the Mantel-Haenszel formula. *Nephron Clin Pract.* 2010;116(4):c317-21.
- Tripepi G, Jager KJ, Dekker FW, Zoccali C. Stratification for confounding-part 2: direct and indirect standardization. *Nephron Clin Pract.* 2010;116(4):c322-5.
- van Stralen KJ, Dekker FW, Zoccali C, Jager KJ. Confounding. *Nephron Clin Pract.* 2010;116(2):c143-7.
- Wunsch H, Linde-Zwirble WT, Angus DC. Methods to adjust for bias and confounding in critical care health services research involving observational data. *J Crit Care.* 2006;21(1):1-7.

## Inferencia causal

La mayor parte de los estudios analíticos en investigación clínica y epidemiológica tiene por objetivo evaluar la existencia de una potencial relación causa-efecto entre el factor de estudio y la variable de respuesta, excluyendo otras posibles explicaciones alternativas de los resultados observados.

Para poder establecer una relación causal es necesario que el estudio tenga validez interna (fig. 31.1). En primer lugar, debe descartarse la existencia de sesgos en la selección de la población o en la información recogida y, si existen, evaluar si pueden llegar a alterar sustancialmente los resultados. También debe valorarse si se han controlado adecuadamente todos los factores de confusión relevantes. Hay que evaluar, además, si en el proceso de análisis se han tenido en cuenta posibles modificaciones del efecto causadas por otras variables. Finalmente, debe valorarse si el resultado final es relevante desde el punto de vista clínico, teniendo en cuenta tanto su magnitud como la precisión con que se ha estimado.

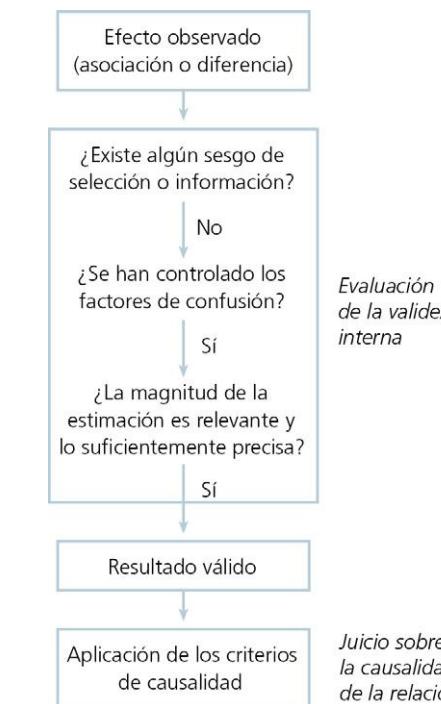
El hecho de que pueda asumirse que el resultado observado es válido no implica que la relación estudiada sea causal. Para evaluar si lo es, hay que tener en cuenta otros aspectos, que suponen un juicio de valor y que, en muchas ocasiones, deben utilizar pruebas y conocimientos procedentes de otros estudios. Por todos estos motivos, difícilmente puede establecerse con seguridad una relación causal a partir de un único estudio.

### TIPOS DE ESTUDIO E INFERENCIA CAUSAL

La validez interna es el grado de solidez o rigor de un estudio, en el sentido de que el modo en que ha sido diseñado y realizado permite estar razonablemente

convencido de que su resultado no está sesgado y proporciona una respuesta correcta a la pregunta planteada. Sin embargo, el estudio perfecto no existe. En cualquier investigación hay multitud de amenazas que pueden comprometer o limitar la validez interna: no suele disponerse de una medida perfecta de la respuesta, los investigadores pueden introducir sesgos o errores inconscientemente, diversos acontecimientos que se producen durante el seguimiento pueden interferir en la medición de la respuesta, pueden existir factores de confusión no conocidos y no controlados, etc. Además, aunque se haya observado un resultado estadísticamente significativo, no se habrá descartado totalmente la posibilidad de que pueda ser debido a la variabilidad aleatoria.

La evaluación de la validez interna está muy relacionada con el diseño utilizado. Si se ha realizado un ensayo clínico aleatorio, bien diseñado y ejecutado, es menos probable que existan sesgos o factores de confusión no controlados que si se ha llevado a cabo un estudio observacional. De hecho, los diseños pueden clasificarse según la calidad de la evidencia que proporcionan sobre una presunta relación causal entre el factor de estudio y la respuesta (fig. 31.2). Este tipo de escalas jerárquicas es muy utilizado a la hora de evaluar las evidencias disponibles relacionadas con una pregunta clínica, de manera que puede establecerse la fuerza de las recomendaciones para la práctica clínica que pueden derivarse. Sin embargo, algunas de estas clasificaciones no discriminan explícitamente entre las posibles diferencias de calidad y rigor que pueden existir entre dos estudios situados en el mismo grado de evidencia científica. Por otro lado, podría darse el caso de que un estudio situado en una categoría inferior presentara mayor calidad y más rigor en



**Figura 31.1** Evaluación de la validez interna como paso previo al juicio sobre la causalidad.



**Figura 31.2** Calidad de la evidencia según el tipo de estudio.

comparación con otro estudio situado en un nivel superior en el esquema de clasificación, pero que no se hubiera ejecutado de forma correcta.

Por ello, la calidad de la evidencia no depende solamente del tipo de diseño utilizado, y no deberían descartarse las pruebas procedentes de estudios

observacionales simplemente porque no corresponden a estudios experimentales. Por lo tanto, la evaluación de la validez interna requiere considerar, además del diseño utilizado, el rigor metodológico del estudio (**cuadro 31.1**).

Obviamente, la definición clara y precisa del objetivo del estudio y la utilización de un diseño de investigación adecuado son imprescindibles para obtener una respuesta válida. También es fundamental la muestra de sujetos que se estudia, no sólo porque debe ser la idónea para la investigación, sino porque será también un factor determinante de la aplicabilidad de los resultados. Además, su tamaño debe ser suficiente para poder detectar diferencias o asociaciones de relevancia clínica estadísticamente significativas.

Cuando un estudio se basa en la comparación de dos o más grupos, es importante que sean concurrentes en el tiempo y similares en los factores pronósticos. Si existen diferencias, deberán tenerse en cuenta en el análisis para evitar que produzcan

#### Cuadro 31.1 Características metodológicas generales asociadas a un mayor rigor científico

1. Definición clara y precisa del objetivo:
  - Factor de estudio y de comparación adecuados
  - Variable de respuesta adecuada y de relevancia clínica
  - Población adecuada
2. Adecuación del diseño a la pregunta de la investigación
3. Muestra adecuada y representativa:
  - Criterios de selección adecuados
  - Procedimiento de selección adecuado
  - Tamaño suficiente
4. Comparabilidad de la línea basal o de inicio en relación con los factores pronósticos (en ensayos clínicos, asignación aleatoria)
5. Comparabilidad a lo largo del estudio:
  - 5.1 Seguimiento de los sujetos:
    - Duración suficiente
    - Completo (reducida proporción de pérdidas)
    - Igual en todos los grupos
  - 5.2 Medición de las variables:
    - Mediciones de suficiente calidad
    - Igual en todos los grupos
    - Enmascaramiento
6. Análisis de datos adecuado:
  - Ajuste por factores de confusión
  - Tratamiento adecuado de las pérdidas y exclusiones
  - Evaluación de un gradiente del tipo dosis-respuesta

un fenómeno de confusión. Dado que la asignación aleatoria tiende a asegurar la comparabilidad en la línea basal, debe utilizarse cuando es factible. Por otro lado, la comparabilidad debe mantenerse a lo largo de todo el estudio, de forma que todos los sujetos deben ser seguidos y estudiados de la misma manera independientemente del grupo al que pertenecen, sometidos a la misma pauta de visitas e intervenciones, y utilizando técnicas de enmascaramiento si es necesario para prevenir sesgos en las mediciones.

Además, el resultado debe haber sido medido por una variable adecuada y el seguimiento debe haber sido lo suficientemente largo y completo para poder detectarlo. Debe tenerse especial prudencia cuando se utilizan variables intermedias o subrogadas como sustitutas de la verdadera respuesta de interés clínico.

El análisis de los datos debe haberse realizado de forma adecuada y la conclusión debe corresponderse con los resultados.

Algunas clasificaciones de la calidad de la evidencia incorporan el metaanálisis como un tipo de estudio, en ocasiones situándolo como el de mayor rigor científico. Sin embargo, a pesar de sus ventajas, no está exento de sesgos ni limitaciones (sesgo de publicación, sesgo de idioma, selección inadecuada de estudios, evaluación sesgada de la calidad de los artículos, etc.) y sus resultados dependen en gran medida de la calidad de los estudios que incluye. Así, por ejemplo, en el caso de que éstos fueran muy heterogéneos o de baja calidad, probablemente la fuerza de la evidencia proporcionada por un único ensayo clínico aleatorio bien diseñado y de gran tamaño podría ser mayor que la del metaanálisis.

**Ejemplo 31.1.** Un estudio tenía por objetivo evaluar las discrepancias entre los resultados obtenidos en ensayos clínicos aleatorios de gran tamaño (más de 1.000 pacientes) y en metaanálisis sobre el mismo tema ([LeLorier et al, 1997](#)). Los autores consideraron los ensayos clínicos como el estándar de referencia. Se incluyeron 12 ensayos clínicos de gran tamaño publicados en cuatro revistas, y 19 metaanálisis que abordaban el mismo problema, y evaluaron los resultados en relación con 40 variables de respuesta, utilizadas en estos estudios. Observaron una concordancia moderada ( $\kappa = 0,35$ ; intervalo de confianza del 95% [IC 95%]: de 0,06 a 0,64). El valor predictivo positivo de los metaanálisis fue del 68% y el negativo del 67%. Es decir, si no se hubieran realizado los ensayos clínicos aleatorios, a partir de los resultados de los metaanálisis, se habrían adoptado tratamientos ineficaces en un 32% de los casos y rechazado tratamientos útiles en el 33%.

## CRITERIOS DE CAUSALIDAD

El establecimiento de una relación causa-efecto es, en gran medida, un juicio de valor que se realiza a partir de las evidencias disponibles. Dado que estas evidencias provienen habitualmente de diferentes estudios, con metodologías distintas, a menudo centrados en aspectos parciales de la cadena causal, a veces proporcionando resultados contradictorios, emitir este juicio no es una tarea fácil, y diferentes investigadores pueden defender opiniones o visiones distintas.

Existen, sin embargo, determinados criterios que pueden ser de ayuda a la hora de distinguir entre asociación y causa. Ninguno de ellos debe entenderse como necesario ni suficiente para establecer una relación causal, sino simplemente como aspectos que hay que tener en cuenta a la hora de realizar esta valoración.

## Magnitud observada de la fuerza de la asociación

Cuanto mayor sea la magnitud de la medida de asociación utilizada (riesgo relativo [RR], *odds ratio*, etc.), mayor será la probabilidad de que el factor de estudio sea la causa de la respuesta observada.

Cuando en un estudio se observa una asociación muy fuerte, es difícil que pueda deberse, en toda su magnitud, a un sesgo o a factores de confusión no controlados. En cambio, si es de escasa magnitud, puede ocurrir que, en realidad, exista una asociación fuerte, pero que un sesgo o un factor no controlado la haga aparecer como débil; pero también puede suceder que el sesgo o los factores no controlados hagan aparecer como positiva una asociación que en realidad no existe, o incluso que es de signo contrario.

La fuerza de una asociación no se mide por el grado de significación estadística. El valor de  $p$  depende de la magnitud de la asociación y del tamaño de la muestra, de forma que asociaciones de escasa o nula relevancia pueden aparecer como estadísticamente significativas en estudios de gran tamaño.

Por otro lado, esta valoración no debe hacerse solamente según la estimación puntual de la magnitud de la asociación, sino que debe tenerse en cuenta la precisión con que se ha valorado, es decir, su IC. Es frecuente encontrar estudios realizados con un número reducido de sujetos, que presentan RR cuyo IC tiene un límite superior muy elevado. Hay que interpretar con cautela estos resultados, ya que reflejan más una ausencia de precisión que una asociación muy fuerte.

**Ejemplo 31.2.** Supongamos un estudio que presenta el siguiente resultado: RR = 19 (IC 95%: 1,5-452). Estos datos no deben interpretarse como que el riesgo

asociado al factor de estudio puede ser hasta 452 veces superior al del grupo control, sino que reflejan una baja precisión en la estimación, probablemente debida a un número reducido de sujetos estudiados. De hecho, lo más cauteloso es interpretar el límite inferior, que indica que, como mínimo, el riesgo se incrementa 1,5 veces.

### Plausibilidad biológica

A la hora de enjuiciar una posible relación causal, el criterio de plausibilidad biológica debería recibir una mayor atención de la habitual. Según él, los resultados obtenidos deben ser compatibles con los conocimientos existentes.

Este criterio se ha utilizado de formas diferentes por los investigadores. Para algunos, una asociación biológicamente plausible es aquella para la cual se puede elaborar una hipótesis razonable sobre su mecanismo, aunque no exista ninguna evidencia biológica sobre él. Argumentan que, si no se adoptara esta visión tan permisiva, podrían descartarse asociaciones causales reales por la ausencia de conocimientos sobre su mecanismo biológico. En realidad, la ausencia de una explicación plausible puede impulsar el desarrollo de nuevas teorías que deben ser sometidas a prueba. En muchas ocasiones, estas hipótesis se generan por analogía con procesos similares. De todas formas, hay que ser cauteloso, ya que puede ser relativamente fácil establecer hipótesis en apariencia biológicamente plausibles a partir de los resultados de un estudio. Además, el uso de paquetes estadísticos informatizados permite relacionar largas listas de variables, por lo que, si el análisis de los datos se hace sin seguir una estrategia previa, aumenta la probabilidad de encontrar asociaciones que carecen de una mínima plausibilidad biológica.

**Ejemplo 31.3.** En un estudio de cohortes en el que se incluyeron prostitutas con la finalidad de analizar factores asociados a la transmisión del VIH entre hombres y mujeres heterosexuales, se observó una fuerte asociación entre el consumo de anticonceptivos orales (AO) y la infección por el VIH. Los autores formularon la hipótesis de que, entre otros mecanismos, el riesgo de transmisión podía estar aumentado, debido al efecto de los AO sobre la mucosa genital, al incrementar el área de ectopia y el riesgo de rotura durante el acto sexual (Plummer et al., 1991). En otro estudio publicado al mismo tiempo, sus autores encontraron un resultado diametralmente opuesto, que indicaba que el uso de AO protegía frente a la infección por el VIH, y formularon otra hipótesis igualmente plausible: dado que los AO,

que contienen progesterona, engrosan la mucosa cervical, dificultarán la entrada del VIH (Lazzarin et al., 1991). Es probable que sesgos y factores de confusión no controlados hayan desempeñado un papel importante en la obtención de estos resultados contradictorios.

Algunos autores argumentan que no es suficiente sugerir simplemente un mecanismo aparentemente plausible, sino que deben existir algunas evidencias que lo apoyen, mientras que otros exigen la presencia de evidencias suficientes que muestren cómo el factor de estudio actúa sobre un mecanismo conocido de producción de la enfermedad. El riesgo de este criterio más restringido puede llevar a descartar hipótesis sólo por la ausencia de conocimientos en un momento determinado.

**Ejemplo 31.4.** En 1977, Winckelstein sugirió una posible asociación entre el consumo de tabaco y el cáncer de cérvix. Propuso dos hipótesis: primero, que el cáncer de cérvix es una enfermedad de células escamosas y que el tabaco causa cánceres de este tipo celular en otros órganos, como el pulmón; y segundo, que los componentes carcinógenos del tabaco podrían ser transportados al epitelio cervical a través de la circulación. No citó ninguna evidencia para estas hipótesis.

Si se hubiera adoptado una postura restrictiva, estas hipótesis hubieran sido descartadas por falta de pruebas. Sin embargo, posteriormente, un estudio presentó resultados sobre concentraciones de nicotina y cotinina en el moco cervical de mujeres fumadoras, lo que representaba los primeros datos que apoyaban las hipótesis anteriores, que empezaron a ganar credibilidad (Sasson et al., 1985). Estudios posteriores han aportado nuevos datos y generado nuevas hipótesis sobre la relación entre el consumo de tabaco y el cáncer de cérvix.

### Consistencia

Según este criterio, los distintos estudios que han abordado un objetivo similar deberían mostrar resultados concordantes. Si varios estudios están realizados en lugares y poblaciones diferentes, y todos llegan a la misma conclusión, se refuerza la idea de causalidad. Sin embargo, es frecuente que diferentes estudios den resultados distintos. Esto no significa que la asociación no sea causal, sino que deben tenerse en cuenta las diferencias en la metodología empleada y la calidad de cada uno de ellos, que podrían explicar los resultados en apariencia contradictorios.

**Ejemplo 31.5.** Los estudios diseñados para evaluar si el consumo de AO causa cáncer de mama han producido resultados contradictorios. Algunos concluyen que las mujeres menores de 35 años expuestas a los AO durante largos períodos tienen un riesgo aumentado de padecer cáncer de mama, mientras que otros no apoyan estos resultados. La revisión cuidadosa de los estudios revela diferencias en los criterios de selección, en las técnicas empleadas para medir la exposición y en otros aspectos del diseño, que podrían explicar las aparentes contradicciones de los resultados.

### Relación temporal

Según este criterio, la presunta causa ha de preceder al efecto y el tiempo de inducción o latencia (tiempo transcurrido entre el comienzo de la exposición y la aparición de la respuesta) debe ser compatible con el que se conoce de la historia natural de la enfermedad. Aunque este principio es evidente, en los estudios transversales y en los de casos y controles con casos prevalentes esta condición es difícil de asegurar.

### Gradiente biológico

El hecho de que la magnitud de la respuesta aumenta a medida que se incrementa el grado de exposición al factor de estudio (en cantidad y/o tiempo) es un argumento más a favor de que la relación entre ambos es causal. Para evaluar este criterio, la exposición al factor de estudio debe tener más de dos categorías, y cada una de ellas debe reflejar los incrementos de la exposición; se trata de determinar

si implican también aumentos del riesgo. Algunos autores incluyen este criterio como un elemento de la plausibilidad biológica.

Sin embargo, no todas las relaciones causales deben tener necesariamente un gradiente dosis-respuesta. Puede ocurrir, por ejemplo, que exista un umbral por encima del cual aumente el riesgo, sin hacerlo más aunque aumente el grado de exposición. Así mismo, la relación no tiene por qué ser lineal; se puede adoptar otro tipo de relación que no se ponga de manifiesto con las técnicas de análisis habituales.

### Disminución del riesgo después de la reducción de la exposición

Como criterio de causalidad, es tan importante observar que al aumentar la dosis de exposición al factor de estudio aumenta el riesgo de enfermedad, como el proceso inverso, es decir, que al reducirse, disminuye el riesgo.

**Ejemplo 31.6.** Desde la década de 1950, diversos estudios de cohortes han demostrado que las personas con concentraciones elevadas de colesterol sérico tienen un mayor riesgo de presentar problemas cardiovasculares. A partir de la década de 1980 se han publicado los resultados de diferentes ensayos clínicos aleatorios que evalúan la eficacia de una intervención para disminuir las cifras de colesterol, y se ha observado que aquellas personas que habían recibido la intervención tenían, en promedio, cifras menores de colesterol sérico que las del grupo control, y también una menor frecuencia de episodios cardiovasculares.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

Lazzarin A, Saracco A, Musicco M, Bicolosi A. Man-to-woman sexual transmission of the human immunodeficiency virus. *Arch Intern Med.* 1991;151:2411-6.  
LeLorier J, Grégoire G, Denhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analysis and

subsequent large randomized, controlled trials. *N Engl J Med.* 1997;337:536-42.  
Plummer FA, Simonsen JN, Cameron DW, Ndinya-Achola JO, Kreiss JK, Gakinya MN, et al. Cofactors in male-female sexual transmission of human immunodeficiency virus type 1. *J Infect Dis.* 1991;163:233-9.

Sasson IM, Haley NJ, Hoffmann D, Wynder EL, Hellberg D, Nilsson S. Cigarette smoking and neoplasia of the uterine cervix: smoke constituents in cervical mucus. *N Engl J Med.* 1985;312:315-6.  
Winckelstein Jr W. Smoking and cancer of the uterine cervix: hypothesis. *Am J Epidemiol.* 1977;106:257-9.

## BIBLIOGRAFÍA

- Argimon JM. Distinguendo entre asociación y causa. FMC. 1998;5:555-6.
- De Vree L. Epidemiology and causation. *Med Health Care Philos.* 2009;12:345-53.
- Goodman C. Literature searching and evidence interpretation for assessing health care practices. Estocolmo: The Swedish Council on Technology Assessment in Health Care; 1993.
- Greenland S, Brumback B. An overview of relations among causal modelling methods. *Int J Epidemiol.* 2002;31:1030-7.
- Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet.* 2002;359:248-52.
- Hennekens CH, DeMets D. Statistical association and causation: contributions of different types of evidence. *JAMA.* 2011;305:1134-5.
- Hernán MA. A definition of causal effect for epidemiological research. *J Epidemiol Community Health.* 2004;58:265-71.
- Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health.* 2006;60:578-86.
- Joffe M. The gap between evidence discovery and actual causal relationships. *Prev Med.* 2011;53:246-9.
- Kaufman JS, Kaufman S. Estimating causal effects. *Int J Epidemiol.* 2002;31:431-2.
- MacLure M, Schneeweiss S. Causation of bias: the episcope. *Epidemiology.* 2001;12:114-22.
- Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol.* 2002;31:422-9.
- Olsen J. What characterises a useful concept of causation in epidemiology? *J Epidemiol Community Health.* 2003;57:86-8.
- Parascandola M. Causes, risks, and probabilities: probabilistic concepts of causation in chronic disease epidemiology. *Prev Med.* 2011;53:232-4.
- Poole C. Causal values. *Epidemiology.* 2001;12:139-41.
- Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health.* 2005;95(Suppl1):S144-50.
- Shafer G. Estimating causal effects. *Int J Epidemiol.* 2002;31:434-5.
- Shapiro S. Causation, bias and confounding: a hitchhiker's guide to the epidemiological galaxy Part 2. Principles of causality in epidemiological research: confounding, effect modification and strength of association. *J Fam Plann Reprod Health Care.* 2008;34:185-90.
- Swaen G, van Amelsvoort L. A weight of evidence approach to causal inference. *J Clin Epidemiol.* 2009;62:270-7.
- Ward A. Causal criteria and the problem of complex causation. *Med Health Care Philos.* 2009;12:333-43.
- Weed DL, Hursting SD. Biologic plausibility in causal inference: current methods and practice. *Am J Epidemiol.* 1998;147:415-25.
- Weed DL. Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related. *Int J Epidemiol.* 2000;29:387-90.
- Weiss NS. Can the «specificity» of an association be rehabilitated as a basis for supporting a causal hypothesis? *Epidemiology.* 2002;13:6-8.
- Worrall J. Causality in medicine: getting back to the Hill top. *Prev Med.* 2011;53:235-8.

## Aplicabilidad práctica de los resultados

La interpretación de los resultados de un estudio incluye tanto la evaluación de la validez interna comentada en los capítulos anteriores, como la de los aspectos relacionados con su aplicabilidad práctica e impacto potencial. Dado que los investigadores son los profesionales que mejor conocen los puntos fuertes y débiles del estudio, están en la situación óptima para poder evaluar su posible utilidad.

La finalidad última de un estudio de investigación es aplicar sus resultados para mejorar el estado de salud de la población, por ejemplo, recomendando los tratamientos más efectivos y seguros, la aplicación de medidas preventivas capaces de disminuir la incidencia de las enfermedades, o el uso de las mejores pruebas diagnósticas. Una vez comprobada la validez interna de los resultados, su aplicabilidad a la práctica clínica se enmarca dentro del concepto de validez externa, lo que supone un juicio de valor para intentar evaluar, por ejemplo, si los resultados pueden ser aplicables a poblaciones de características diferentes a las de la muestra estudiada o a entornos con distintas condiciones. Esta valoración se fundamenta tanto en la magnitud del efecto observado, ya que si es de escasa relevancia es más probable que no se produzca en otras poblaciones, como en su plausibilidad biológica, en el sentido de evaluar si es razonable asumir que el mecanismo fisiopatológico por el que se produce el efecto se cumple también en poblaciones de diferentes características.

### FACTOR DE ESTUDIO Y DE COMPARACIÓN

Para poder aplicar los resultados de un estudio es necesario que la pregunta que se ha formulado tenga relevancia clínica y corresponda a una cuestión de interés para los profesionales.

En un estudio sobre pruebas diagnósticas es importante considerar si es factible que la prueba que se evalúa pueda realizarse en un entorno determinado y si los profesionales tienen las habilidades necesarias para llevarla a cabo e interpretar correctamente sus resultados. Dado que el estudio se ha realizado en condiciones más o menos ideales, con profesionales motivados y entrenados en la aplicación de la prueba, es probable que, cuando sea aplicada en la práctica por personas no tan entrenadas en su manejo, puedan obtenerse resultados diferentes. Además, hay que considerar la información sobre la variabilidad interobservador e intraobservador. Si se trata de un ensayo clínico que evalúa un tratamiento, la descripción de la intervención debe ser lo suficientemente detallada para poder evaluar su aplicabilidad.

**Ejemplo 32.1.** Duff et al (2010) estudiaron 262 ensayos clínicos que evaluaban tratamientos oncológicos y que habían sido publicados en cinco revistas importantes entre 2005 y 2008. Analizaron si los artículos contenían la información necesaria para permitir la aplicación clínica del tratamiento (nombre del fármaco, dosis, vía de administración, duración de los ciclos, número de ciclos, premedicación, parámetros de monitorización, etc.), y encontraron que tan sólo 30 artículos (11%) cumplían los 10 criterios considerados esenciales para poder trasladar los resultados a la práctica clínica.

La aplicabilidad depende también de la alternativa de referencia, en el sentido de si la comparación se realiza entre las opciones que realmente interesan, de forma que el ensayo clínico sirva de ayuda en la toma de decisiones en la práctica clínica.

**Ejemplo 32.2.** La recomendación de la atorvastatina a dosis elevadas (80 mg/día) en pacientes con enfermedad coronaria se basa fundamentalmente en dos ensayos clínicos. En uno de ellos se comparó con una dosis mucho menor (10 mg/día) del mismo fármaco (La Rosa et al, 2005), y en el otro frente a una dosis menos potente de otra estatina (pravastatina, 40 mg/día) (Cannon et al, 2004). Si estos estudios hubieran incluido un grupo con una dosis intermedia de atorvastatina (40 mg/día), hubieran podido demostrar una equivalencia terapéutica entre la dosis elevada y la intermedia, lo que hubiera evitado los costes y el riesgo de toxicidad asociados al uso de elevadas dosis del fármaco en un gran número de pacientes.

Un aspecto controvertido es el uso de placebo como comparación cuando ya existe una alternativa que ha demostrado su eficacia y es utilizada por los profesionales clínicos. Lo que realmente interesa es comparar el nuevo tratamiento con la mejor alternativa existente, y no frente a placebo. Dado que cada vez es más difícil esperar que los nuevos representen importantes incrementos de eficacia, muchos investigadores e industrias farmacéuticas realizan los estudios comparando con placebo, ya que esperan encontrar diferencias de eficacia de mayor magnitud y, por consiguiente, requieren estudios de un tamaño más reducido. Sin embargo, hay que valorar si estas comparaciones vulneran los principios éticos de la investigación en seres humanos.

Hay aspectos de la aplicación del factor de estudio que pueden limitar su extrapolación a la práctica clínica en determinadas condiciones. Por ejemplo, si se trata de una intervención compleja o no farmacológica, cuyos resultados dependen en gran medida de la experiencia y el entrenamiento previos de los profesionales, es importante tomar en consideración estos aspectos a la hora de valorar su aplicabilidad, ya que pueden alterar la relación entre riesgos y beneficios en función del entorno en que deba aplicarse.

**Ejemplo 32.3.** Supongamos un ensayo clínico que compare dos técnicas de criocirugía cutánea (nitrógeno líquido y un aerosol de dimetil-éter-propano) en consultas de atención primaria, en el que no se observan diferencias clínicamente relevantes en cuanto a eficacia, tolerancia y seguridad de ambos agentes criogénicos. La evaluación de la aplicabilidad de los resultados requiere tener en cuenta las ventajas y desventajas de cada una de las técnicas, la capacidad diagnóstica de los profesionales para identificar las lesiones dermatológicas susceptibles de recibir tratamiento, las habilidades técnicas

adquiribles por cualquier profesional con una formación adecuada, así como los requerimientos técnicos necesarios para el adecuado almacenamiento de las sustancias (p. ej., el nitrógeno líquido precisa unos contenedores especiales, habitualmente no disponibles en los centros de salud). Este tipo de estudios suelen hacerlo profesionales interesados y motivados por el tema, formados y con experiencia. Por tanto, hay que tener mucha precaución a la hora de extrapoluar los resultados a ámbitos en los que la criocirugía cutánea no es una práctica habitual en las consultas.

Algo similar ocurre, por ejemplo, con fármacos en los que es necesario monitorizar sus concentraciones plasmáticas, si se decide aplicarlos a entornos en los que la tecnología necesaria no está disponible. O cuando la intervención puede producir efectos adversos graves que requieren un tratamiento sofisticado o que son de difícil diagnóstico, lo que aumentaría la gravedad de las consecuencias de los efectos adversos y podría interferir en la evaluación de la relación entre los riesgos y los beneficios esperados.

## VARIABLE DE RESPUESTA

Para poder elaborar recomendaciones sólidas a partir de los datos de una investigación, idealmente ésta debería evaluar el impacto de un factor de estudio (prueba diagnóstica, tratamiento, actividad preventiva, supresión de una exposición, etc.) sobre una variable de resultado final (mortalidad, incidencia de enfermedades, calidad de vida, etc.). Sin embargo, pocos estudios lo documentan directamente, dadas las dificultades metodológicas de evaluar completamente un proceso en el que influyen múltiples variables difíciles de controlar y que requerirían seguimientos muy prolongados y un número elevado de sujetos. En consecuencia, en muchas ocasiones los estudios se limitan a evaluar aspectos parciales de la relación.

**Ejemplo 32.4.** En la situación representada en la figura 32.1, lo ideal sería realizar un estudio que demuestre si la detección precoz de concentraciones elevadas de colesterol en sangre mejora el pronóstico de los sujetos y así reducir la incidencia y la mortalidad cardiovasculares. Sin embargo, lo que se suele hacer es realizar estudios que aporten evidencias parciales, por ejemplo, sobre la utilidad de diferentes técnicas de detección de las concentraciones elevadas de colesterol sérico o sobre si la aplicación de diferentes intervenciones, ya sean farmacológicas o dietéticas, produce una disminución de las



**Figura 32.1** Detección de las concentraciones elevadas de colesterol sérico: marco para la elaboración de recomendaciones.

concentraciones de colesterol o, lo que sería aún mejor, de la incidencia de la enfermedad cardiovascular. Cuantos más pasos de este proceso abarque un estudio, más firmes podrán ser sus recomendaciones. Así, un estudio que evalúe la eficacia de una intervención dietética sobre la incidencia de enfermedad cardiovascular proporcionará una base más sólida para recomendarla que si tan sólo ha demostrado que reduce las concentraciones de colesterol.

La variable de respuesta utilizada en un estudio es clave para valorar el interés de sus resultados. En el capítulo 13 se han comentado los criterios y las dificultades relacionadas con su elección, así como los problemas del uso de múltiples resultados, de variables subrogadas o de variables compuestas.

Muchos estudios utilizan variables intermedias o subrogadas, asumiendo que se correlacionan con el resultado principal de interés. Los resultados de estos estudios deben interpretarse con cautela, ya que, por ejemplo, el hecho de que un fármaco disminuya el colesterol y los triglicéridos no implica que aumente la supervivencia, el de que un antihipertensivo reduzca la presión arterial no significa que disminuya el riesgo cardiovascular, o el de que un tratamiento aumente la densidad mineral ósea no necesariamente disminuye el riesgo de fracturas. De hecho, muchas variables intermedias pueden ser consideradas válidas durante años, e investigaciones posteriores pueden demostrar que no lo eran.

**Ejemplo 32.5.** Duffy y Treasure (2011) comentan el caso de la fuerte asociación entre la detección de anticuerpos frente al virus de Epstein-Barr y el cáncer nasofaríngeo en población china, con valores de riesgo relativo (RR) entre 30 y 130. La prueba de detección tiene una sensibilidad y una especificidad superiores al 90%, y el 5-10% de los sujetos que presentan un resultado positivo supone el 80-95% de los casos de cáncer. Por lo tanto, la detección de anticuerpos puede ser potencialmente una buena prueba de cribado. También se ha propuesto que la disminución en la respuesta de los anticuerpos podría utilizarse como variable subrogada de la supervivencia en los ensayos clínicos. Para poder hacerlo, es necesario demostrar que dicha disminución supone un beneficio terapéutico real.

A la hora de interpretar los resultados de un estudio que ha utilizado una variable intermedia, hay que valorar también si su relación con la variable principal es causal, si se trata de una asociación coincidente o mediada por un tercer factor, o si tan sólo refleja uno de los mecanismos de actuación sobre la variable principal.

**Ejemplo 32.6.** La relación causa-efecto entre la hipertensión arterial y el aumento de riesgo de mortalidad cardiovascular está bien establecida. Sin embargo, el beneficio del tratamiento antihipertensivo sobre la mortalidad cardiovascular no puede generalizarse directamente de su eficacia sobre una variable intermedia como es la medición de la presión arterial. Algunos fármacos, como los diuréticos o los bloqueadores betaadrenérgicos, han demostrado en ensayos clínicos controlados mejoras de supervivencia en los pacientes del grupo sometido a intervención, mientras que otros grupos de fármacos o no lo han demostrado, o hay datos que sugieren un efecto negativo a través de mecanismos de acción mal conocidos que contrarrestarían el beneficio obtenido por la reducción de la presión arterial (Psaty et al, 1997).

El uso y la confianza excesivos en las variables subrogadas puede llegar a desplazar el objetivo terapéutico. Así, muchas guías establecen dicho objetivo, no en relación con variables finales de interés clínico, sino con variables subrogadas. Como consecuencia, se instauran tratamientos cada vez más intensivos buscando alcanzar dicho objetivo, aunque no hayan demostrado que tengan efecto sobre la variable final de interés, lo que a la larga puede implicar un verdadero riesgo para los pacientes.

Desde un punto de vista práctico, cuando un fármaco ha demostrado su efecto sobre una variable intermedia, y se ha comprobado que se corresponde con una mejora en la variable principal, al evaluar el efecto de nuevos fármacos del mismo grupo terapéutico, no es imprescindible estudiar de nuevo sus efectos sobre la variable principal. Sin embargo, los resultados de estos estudios deben interpretarse con cautela, ya que la asunción de este efecto de clase no siempre es cierta.

**Ejemplo 32.7.** Las pruebas sobre un efecto de clase de los inhibidores de la enzima convertidora de angiotensina (IECA) en los pacientes con enfermedad cardiovascular estable y función ventricular izquierda preservada presentan alguna discordancia. Así, mientras que el estudio HOPE mostró que el ramipril reducía los eventos cardiovasculares en un 18% en relación con placebo y el estudio EUROPA que perindopril lo hacía en un 20%, el estudio PEACE no apreció una reducción de dichos eventos con trandalopril. Se ha intentado explicar esta discordancia por niveles diferentes de penetración tisular, o bien por diferencias en el diseño de los estudios (definición de la variable de respuesta, duración, riesgo basal, cointervenciones, etc.). La demostración de la superioridad o equivalencia de los fármacos de la misma clase debería realizarse a través de ensayos clínicos que los compararan directamente.

De todas formas, el hecho de haber demostrado que un factor de estudio tiene un efecto relevante sobre una única variable principal no es suficiente para elaborar una recomendación sin haber evaluado sus efectos sobre otras variables también importantes. Por ejemplo, es posible que un tratamiento antineoplásico aumente la supervivencia de los pacientes con determinado tipo de cáncer, pero disminuya su calidad de vida; o que una intervención quirúrgica aumente la supervivencia a largo plazo de los pacientes, pero aumente la mortalidad peroperatoria; o que el ácido acetilsalicílico a determinadas dosis sea capaz de prevenir el infarto agudo de miocardio (IAM), pero tenga efectos secundarios importantes (hemorragias, accidentes cerebrovasculares, etc.). La presentación de los resultados de los estudios en las publicaciones debería incluir la información necesaria para evaluar adecuadamente, no sólo la existencia de un efecto estadísticamente significativo, sino también su relevancia en la práctica clínica.

**Ejemplo 32.8.** En una revisión de 27 ensayos clínicos publicados en cinco revistas de prestigio, Chan et al (2001) encontraron que solamente 22 (81%) identificaban explícitamente una única variable de respuesta principal. De los 20 artículos que habían calculado el tamaño de la muestra necesario, 18 (90%) informaban de la diferencia en que basaban dicho cálculo, aunque tan sólo 2 (11%) la habían establecido como la mínima diferencia clínicamente importante. Por otro lado, de los 27 ensayos, solamente 11 (41%) presentaban el intervalo de confianza (IC) del resultado principal y 20 (74%) interpretaban los resultados desde la perspectiva de su relevancia clínica. Concluyen que los autores de

los artículos no siempre interpretan la importancia clínica de sus resultados ni proporcionan la información suficiente para que los lectores puedan realizar su propia interpretación.

## MEDIDAS DE IMPACTO

Los estudios analíticos están diseñados para estimar el efecto de un factor de estudio sobre una variable de respuesta. Esta estimación suele hacerse a través de medidas que estiman la fuerza de la asociación, como el RR o la *odds ratio* (OR) (anexo 2), que deberían complementarse con medidas del impacto potencial sobre la salud de la población.

Desde un punto de vista epidemiológico pueden distinguirse dos tipos de medidas de impacto: diferencia de incidencias (DI) entre los grupos estudiados y proporción de la incidencia en un grupo o población que puede ser atribuida al factor de estudio.

La DI o diferencia de riesgos se define como la diferencia entre la incidencia en el grupo de sujetos expuestos al factor de estudio ( $I_1$ ) y la incidencia en el de no expuestos ( $I_0$ ) (cuadro 32.1), y representa la cantidad de incidencia que puede ser atribuida al factor de estudio. A diferencia del RR, la DI tiene en cuenta la magnitud de las incidencias que se están comparando.

**Ejemplo 32.9.** Cuando se comparan incidencias acumuladas del 90 y del 30%, se obtiene un RR = 3, al igual que si las incidencias fueran del 3 y del 1%. En cambio, la DI informa sobre el exceso de riesgo en los individuos expuestos en relación con los no expuestos, y sería del 60% en el primer caso, y sólo del 2% en el segundo.

Desde el punto de vista de la salud pública, la DI es más informativa, ya que muestra el exceso de casos que están asociados a la exposición al factor de estudio. Por otro lado, desde el punto de vista de la investigación etiológica, el RR es más informativo, ya que estima la fuerza de la asociación entre el factor de estudio y la respuesta. La DI no sólo depende del RR, sino también de la frecuencia de la enfermedad. Para enfermedades poco frecuentes, aun cuando el RR asociado a una exposición sea elevado, la DI puede ser más pequeña que si el RR de esa misma exposición es menor, pero la enfermedad mucho más frecuente.

**Ejemplo 32.10.** El ejemplo más clásico es la asociación entre el consumo de tabaco y el cáncer de

**Cuadro 32.1 Medidas de impacto**

Diferencia de incidencias (DI) (1)

$$DI = I_1 - I_0$$

Riesgo atribuible en los expuestos (RAE):

$$RAE = \frac{(I_1 - I_0)I_1}{I_1} = \frac{RR - 1}{RR}$$

(1) En un ensayo clínico aleatorio, se denominan reducción absoluta de riesgo (RAR)

$$RAR = (I_1 - I_0)I_1$$

$$RAR = (RR - 1)/RR^*$$

Riesgo atribuible poblacional (RAP):

$$RAP = (Ip - I_0)Ip$$

$RAP = RAE \times \text{Proporción de exposición entre los enfermos}^{**}$

Reducción relativa de riesgo que se produce con el tratamiento (RRR)

$$RRR = [Pc - Pt]/Pc$$

$I_1$ : incidencia en el grupo de estudio.

$I_0$ : incidencia en el grupo de referencia.

RR: riesgo relativo.

Ip: incidencia en la población.

Pc: proporción de sujetos que desarrollan la respuesta en el grupo control.

Pt: proporción de sujetos que desarrollan la respuesta en el grupo de intervención.

\* Cuando la medida del efecto es la OR, sustituir RR por OR.

\*\* En una tabla típica de  $2 \times 2$ , la proporción de exposición entre los enfermos es el cociente  $a/(a+c)$ .

pulmón, por un lado, y la cardiopatía isquémica, por otro. El RR entre el consumo de tabaco y el cáncer de pulmón es mayor de 10, mientras que su asociación con la cardiopatía isquémica es aproximadamente de 2, pero al ser esta última enfermedad mucho más frecuente, su DI es mayor. Esto significa que el consumo de tabaco está asociado con mayor fuerza al cáncer de pulmón, pero causa más casos de IAM que de cáncer.

El riesgo atribuible proporcional, o fracción atribuible, representa la DI relativa. Existen dos medidas de riesgo atribuible proporcional, según la incidencia que se utilice de referencia para el cálculo (ver cuadro 32.1). Si el denominador es la incidencia en el grupo expuesto al factor de riesgo, se denomina *riesgo atribuible en los expuestos* (RAE). Si el denominador es la incidencia en la población total ( $I_p$ ),

compuesta tanto por sujetos expuestos como no expuestos, se denomina *riesgo atribuible poblacional* (RAP). El RAE y el RAP se suelen expresar como porcentajes y se interpretan como la proporción de la incidencia en el grupo de referencia (el utilizado en el denominador) que puede ser atribuida al factor de estudio. También pueden calcularse a partir del RR y de la OR (ver cuadro 32.1).

Las medidas de impacto permiten conocer la magnitud del daño ocasionado por un factor de riesgo determinado y dan una idea del impacto que tendría sobre la incidencia de una enfermedad la eliminación de dicho factor, por lo que se han usado en planificación sanitaria para priorizar intervenciones. Una de sus principales limitaciones se relaciona con su propia interpretación. Así, por ejemplo, el RAP se interpreta como la proporción en la que se reduciría la incidencia de la enfermedad si se eliminase totalmente el factor de riesgo. Plantearse la eliminación total de un factor de riesgo no suele ser realista, por lo que estas medidas son una estimación del impacto potencial, pero no del real.

**Ejemplo 32.11.** En el estudio de Framingham se estimaron los RR de padecer enfermedad coronaria asociados a la presencia de hipertensión arterial (HTA), obesidad y diabetes, que fueron de 2,2, 1,3 y 2,1, respectivamente. Sin embargo, el RAP fue, respectivamente, del 16,7, el 7,8 y el 0,7% para cada uno de los tres factores de riesgo. Estas cifras indican que, a pesar de que la diabetes tenga un RR similar al de la HTA y claramente superior al de la obesidad, su RAP es bastante inferior, dado que su prevalencia es menor que la de los otros factores. Esto indica que un programa dirigido a reducir la incidencia de la enfermedad coronaria mediante el control de la diabetes tendría menos impacto que otro enfocado al control de la HTA o de la obesidad, siempre que las condiciones de factibilidad, relevancia social y eficiencia de los programas fueran similares.

Otra situación es cuando se mide el impacto de un mismo factor de riesgo en la incidencia de distintas enfermedades.

**Ejemplo 32.12.** En el estudio de Framingham se ha estimado que el RR asociado a la HTA en relación con la enfermedad coronaria es de 2,2, mientras que el RR asociado a la enfermedad vascular periférica es de 2,5. Dado que los RR son similares para ambas enfermedades, el RAE y el RAP también lo serán, ya que se trata de un mismo factor de riesgo (de hecho, el RAE para la enfermedad coronaria es del 57% y para la enfermedad vascular periférica, del 60%). En esta situación, la incidencia de la enfermedad

desempeña un papel crucial, por lo que es necesario recurrir a la DI para obtener una visión diferente del impacto del factor de riesgo. La DI para la enfermedad coronaria es del 14,2% a los 10 años, mientras que para la enfermedad vascular periférica es sólo del 3,7%. La DI es superior en el caso de la enfermedad coronaria, dada su mayor incidencia.

Cuando el estudio corresponde a un ensayo clínico aleatorio que evalúa el efecto de un tratamiento, la DI se denomina *reducción absoluta de riesgo* (RAR) y el RAE se conoce como reducción relativa de riesgo (RRR). Otra medida útil para expresar el efecto de un tratamiento es el número de casos que es necesario tratar para prevenir un suceso (NNT). Las medidas que permiten cuantificar el efecto de un tratamiento se comentan en el anexo 5.

## POBLACIÓN DE ESTUDIO

Desde un punto de vista estricto, los hallazgos de un estudio sólo son aplicables a los individuos similares a los que han participado en el mismo. Sin embargo, su extrapolación a otros pacientes es de gran importancia para los clínicos. El primer paso para realizar esta inferencia es evaluar si los participantes son semejantes a los pacientes a los que se espera aplicar el tratamiento o evitar la exposición. Una de las principales críticas al diseño de la mayoría de los ensayos clínicos es que habitualmente excluyen determinados grupos de población, lo que limita la validez externa de los resultados.

**Ejemplo 32.13.** Al revisar los criterios de selección de participantes en 283 ensayos clínicos publicados en revistas con un elevado factor de impacto, Van Spall et al (2007) observaron que se excluyeron pacientes por la edad en el 72% de los ensayos (por edad pediátrica en el 60% y por ancianos en el 38,5%), por condiciones relacionadas con el sexo femenino en el 39%, por problemas médicos comunes en el 81% y por tomar con frecuencia medicamentos prescritos en el 54%. Los criterios de exclusión no estaban suficientemente justificados en el contexto específico del ensayo en el 53% de los casos. Estas exclusiones eran más frecuentes en los ensayos multicéntricos, los que evaluaban intervenciones farmacológicas y los promovidos por la industria farmacéutica.

**Ejemplo 32.14.** Travers et al (2007) estudiaron la proporción de individuos con asma bronquial que hubieran sido elegibles en los 17 principales ensayos clínicos citados en las recomendaciones de la Global Initiative for Asthma (GINA). Para ello, utilizaron

los datos de una encuesta poblacional sobre salud respiratoria realizada en una muestra aleatoria de una comunidad. Solamente cumplían los criterios de selección de los ensayos entre el 0 y el 36% de los pacientes con asma (mediana del 4%) y entre el 0 y el 43% de los que estaban en tratamiento (mediana del 6%). Así pues, los ensayos en los que se basan las recomendaciones terapéuticas tienen una reducida validez externa al haberse realizado en muestras de sujetos altamente seleccionadas.

Si los criterios de selección de los participantes son muy restrictivos (actitud explicativa), se consigue una muestra de sujetos muy homogénea, lo que favorece la validez interna; además, se espera que el efecto sea frecuente, de modo que se gana en eficiencia, ya que se requieren menos individuos y/o menos tiempo de seguimiento, a la vez que aumenta la posibilidad de obtener resultados estadísticamente significativos. Sin embargo, la extrapolación a sujetos de características diferentes a las estudiadas puede presentar limitaciones importantes. Por otro lado, si se han utilizado criterios poco restrictivos (actitud pragmática), se tiende a obtener muestras de sujetos de características más similares a las de los sujetos visitados en la práctica habitual y, por tanto, los resultados son más fácilmente extrapolables, pero con el inconveniente de que la población estudiada es más heterogénea y, por tanto, es más difícil detectar un efecto, a la vez que se requiere un mayor número de sujetos y/o tiempo de seguimiento para contrarrestar esta mayor variabilidad.

Cuando se desea elaborar recomendaciones a partir de los resultados de un estudio, hay que tener en cuenta las diferencias entre las características de los sujetos estudiados y las de aquellos que conforman la población diana de la recomendación, ya que las diferencias entre ellas podrían dar lugar a modificaciones en la magnitud de la respuesta. La extrapolación de los resultados, como última fase de la interpretación, tampoco es una tarea mecánica que se derive simplemente de un solo estudio o de una prueba estadística, sino que requiere un juicio de valor para el que la existencia de mecanismos fisiológicos y patológicos que puedan explicarlos son de gran ayuda.

**Ejemplo 32.15.** El American Physicians Health Study (1988) es un ensayo clínico aleatorio, doble ciego y controlado con placebo. Uno de sus objetivos era evaluar si 325 mg de ácido acetilsalicílico a días alternos reducía la incidencia de IAM. La población potencialmente candidata eran todos los hombres médicos de edades comprendidas entre los 40 y los 84 años que residían en Estados Unidos al inicio

del estudio. Se enviaron cartas a 261.248 médicos solicitando su colaboración; de estos, respondieron 112.528, de los cuales 59.285 aceptaron participar en el estudio. Los principales criterios de exclusión fueron los antecedentes personales de enfermedad cardiovascular, cáncer, enfermedad renal o hepática, úlcera péptica ogota, contraindicaciones de uso de aspirina, consumo de aspirina u otros agentes antiplaquetarios y consumo de antiinflamatorios no esteroideos.

Un total de 33.223 médicos cumplían los criterios de selección y fueron sometidos a un período de preinclusión de 18 semanas en el que tomaron aspirina con el propósito de excluir, antes de la asignación aleatoria, a los sujetos no cumplidores o que desarrollaran efectos secundarios. Finalmente, 22.071 médicos fueron asignados aleatoriamente a los grupos de estudio: 11.037 al grupo que recibía aspirina y 11.034 al grupo placebo. Después de un seguimiento promedio de 5 años, se habían producido 139 casos de IAM (10 de ellos mortales) en el grupo de tratamiento activo y 239 (26 de ellos mortales) en el grupo placebo. El RR era de 0,56 (IC 95%: 0,45 a 0,70) para el total de IAM y de 0,34 (IC 95%: 0,15 a 0,75) para los IAM mortales.

Supongamos que debemos decidir si estos hallazgos son aplicables a las mujeres no médicos que cumplen el resto de los criterios de selección. Para abordar esta cuestión hay que considerar, en primer lugar, si el mecanismo de acción de la aspirina en la prevención del IAM es aplicable a las mujeres no médicas, es decir, si existe alguna diferencia en la fisiopatología de la enfermedad que haga sospechar que los resultados no son extrapolables. Sin embargo, en muchas ocasiones no es posible llegar a una conclusión de esta naturaleza de forma inequívoca, dado que nuestro conocimiento sobre la patogenia de las enfermedades no es definitivo. Mientras no exista ningún dato que haga presumible que los no médicos no puedan beneficiarse del tratamiento, en las mujeres hay que considerar las diferencias hormonales y otras, que podrían resultar en una extrapolación inexacta. Esta incertidumbre sería aún mayor si se decidiera generalizar los datos a hombres y/o mujeres que no cumplen los criterios de inclusión del estudio, o bien extraerlos si con una dosis menor se hubiera podido conseguir el mismo beneficio.

Muchos autores han cuestionado hasta qué punto los resultados de un estudio pueden ser aplicables a pacientes que no han sido incluidos en él, ya que los criterios de inclusión suelen restringir la entrada a pacientes con menor comorbilidad y mejor pronóstico, se realizan en centros sanitarios

de excelencia con profesionales experimentados y mejores equipamientos, los pacientes siguen un protocolo más riguroso, probablemente con un mejor cumplimiento de la intervención, y han dado su consentimiento para participar, lo que indica que adquieren cierto grado de compromiso para realizar un mejor cumplimiento de todas las recomendaciones e intervenciones que se les practican, etc. Por tanto, no es sorprendente que, en promedio, los pacientes incluidos en un ensayo clínico tengan un mejor pronóstico que los no incluidos.

**Ejemplo 32.16.** Steg et al (2007) evaluaron a 8.469 pacientes consecutivos con IAM incluidos en el Global Registry of Acute Coronary Events (GRACE), y los clasificaron en tres grupos: participantes en un ensayo clínico ( $n=953$ , 11%), elegibles para su participación pero no incluidos ( $n=4.669$ , 55%) y no elegibles ( $n=2.847$ , 34%). El riesgo de defunción era menor en el primer grupo, mayor en el segundo y mucho mayor en el tercero. La mortalidad hospitalaria presentó un patrón similar: del 3,6, el 7,1 y el 11,4%, respectivamente. El ajuste por el riesgo basal, la realización de reperfusión y el tipo de la misma, y el retraso en la presentación, mostraron que las diferencias no se explicaban por estas características. Así pues, los pacientes incluidos en el ensayo clínico tenían un riesgo basal y experimentaban una mortalidad menor que los no incluidos, incluso aunque fueran elegibles para el estudio.

## Subgrupos de población

Habitualmente, los resultados se presentan de forma global para todos los sujetos estudiados. Sin embargo, cuando las poblaciones son heterogéneas, el análisis de subgrupos puede resultar de gran utilidad para comparar subgrupos de sujetos definidos, por ejemplo, por diferencias en su riesgo basal. De este modo, un tratamiento que ha mostrado un efecto beneficioso en el conjunto de los individuos estudiados, puede no serlo en todos por igual, e incluso puede ser perjudicial en algunos de ellos.

**Ejemplo 32.17.** Rothwell (1995) reanalizó los resultados del European Carotid Surgery Trial (ECST) para ilustrar este punto. En dicho estudio se comparaba el tratamiento quirúrgico (endarterectomía carotídea) frente al no quirúrgico en pacientes con accidentes cerebrales isquémicos transitorios (AIT).

El reanálisis se circunscribió a los pacientes con estenosis de carótida grave, que fue en quienes se observó un beneficio del tratamiento quirúrgico.

Este autor obtuvo, mediante la aplicación de un modelo de regresión, un índice pronóstico de padecer un AIT de todos los participantes en el estudio. Los individuos incluidos en los grupos de intervención y control fueron clasificados en tres categorías de riesgo: bajo, intermedio y alto. A continuación, calculó los riesgos absolutos y relativos del tratamiento en cada una de las tres categorías (tabla 32.1). Los resultados muestran que, en global, el tratamiento quirúrgico es beneficioso en toda la muestra, y en particular, en los de riesgo alto e intermedio. Estos resultados son estadísticamente significativos, ya que el IC de la RRR excluye el valor 1. Sin embargo, en el subgrupo de riesgo bajo el resultado no es estadísticamente significativo, y el IC muestra que la intervención incluso podría ser perjudicial en este grupo.

Los análisis de subgrupos suelen considerarse sospechosos y poco fiables, lo que es cierto cuando no se han realizado adecuadamente. Sin embargo, cuando se han planificado a priori y se basan en hipótesis bien fundamentadas, pueden ser útiles para poner de manifiesto la existencia de diferencias entre subgrupos de sujetos. Por ello, debe evaluarse si existe heterogeneidad o interacción (modificación de efecto) entre los subgrupos. Si no existe, puede utilizarse la estimación relativa del efecto que se ha observado en el conjunto de los pacientes. Si existe, deben utilizarse estimaciones separadas para cada subgrupo.

**Ejemplo 32.18.** Hernández et al (2006) revisaron 63 ensayos clínicos de más de 100 pacientes del área cardiovascular publicados en 2002 y 2004. De los 39 (62%) que presentaban análisis de subgrupos, ninguno se había diseñado específicamente para tener la potencia suficiente para dicho análisis, solamente en 14 los subgrupos habían sido especificados a priori y en 11 se habían realizado pruebas de interacción. En 21 casos se presentaban conclusiones referidas a resultados estadísticamente significativos en subgrupos y en 15 de ellos con mayor énfasis que en los resultados globales del ensayo. Los autores concluyen que es necesario mejorar la información sobre los análisis de subgrupos, ya que el énfasis sobre sus resultados puede conducir a decisiones terapéuticas erróneas.

### Riesgo basal de la población

Cuando se desea aplicar el resultado de un estudio a un paciente concreto, debe tenerse en cuenta que, de hecho, representa un promedio de los efectos observados en el conjunto de sujetos estudiados y que, si la muestra es heterogénea, puede no reflejarlos adecuadamente. Así, aunque el paciente concreto cumpla los criterios de selección del estudio, es posible que el resultado promedio no estime correctamente el efecto esperable en dicho sujeto. El análisis de subgrupos, si se ha realizado rigurosamente, puede proporcionar

**Tabla 32.1** Análisis de subgrupos planteado en el ejemplo 32.13

	Grupo control			Grupo intervención		
	n	ACV	(riesgo)	n	ACV	(riesgo)
<b>Riesgo</b>						
Bajo	119	10	(8,4%)	183	18	(9,8%)
Intermedio	178	25	(14,0%)	273	19	(7,0%)
Alto	89	18	(20,2%)	132	8	(6,1%)
Total	386	53	(13,7%)	588	45	(7,7%)
<b>Reducción relativa del riesgo (IC 95%)</b>				<b>Reducción absoluta del riesgo</b>		
<b>Riesgo</b>						
Bajo	1,16		(0,6-2,4)			-1,4%
Intermedio	0,53		(0,3-0,9)			7,0%
Alto	0,34		(0,1-0,7)			14,1%
Total	0,51		(0,3-0,8)			6,0%

**Tabla 32.2** Número de personas que deben ser tratadas para prevenir un episodio en función del riesgo basal en el grupo de referencia (ejemplo 32.14)

	Riesgo basal (grupo de referencia)		
	20%	10%	1%
Riesgo en el grupo tratado	14%	7%	0,7%
Reducción relativa de riesgo (RRR)	30%	30%	30%
Reducción absoluta de riesgo (RAR)	6%	3%	0,3%
Número de personas que deben ser tratadas para prevenir un episodio (NNT)	$1/0,06 = 16,7$	$1/0,03 = 33,3$	$1/0,003 = 333,3$

**Tabla 32.3** Efecto de la prevalencia sobre los valores predictivos de una prueba diagnóstica (se asume una sensibilidad y una especificidad del 95%)

Prevalencia %	50	10	1
Valor predictivo de una prueba positiva %	95	68	16
Valor predictivo de una prueba negativa %	95	99,4	99,9

estimaciones del efecto en grupos más homogéneos de sujetos, pero a menudo no es suficiente para tomar la decisión clínica adecuada.

Las medidas de impacto que cuantifican el efecto de un tratamiento (anexo 5) varían en función del riesgo basal del sujeto. En la tabla 32.2 se presenta el ejemplo de un tratamiento que produce una RRR del 30%, y muestra cómo, a medida que disminuye el riesgo basal, aumenta el esfuerzo terapéutico necesario para prevenir un episodio. El problema a la hora de aplicar los resultados a la práctica es disponer de información sobre el riesgo basal de los sujetos.

Este mismo concepto es aplicable a cualquier tipo de estudio. Si una investigación tiene por objetivo obtener una ecuación predictiva del riesgo de presentar una enfermedad, su aplicabilidad a po-

blaciones de características diferentes de la estudiada debe tener en cuenta el entorno y las características de los sujetos en los que se ha determinado la ecuación. Esto es lo que ocurre con muchas de las reglas de cálculo de riesgo de padecer un episodio cardiovascular que se han derivado de poblaciones anglosajonas que presentan una incidencia basal y una mortalidad muy diferentes a las de la población española.

De forma similar, en los estudios que evalúan una prueba diagnóstica puede suceder que su sensibilidad y especificidad puedan ser diferentes en subgrupos determinados de pacientes. Además, los valores predictivos que miden la utilidad práctica de la prueba varían en función de la prevalencia de la enfermedad en la población en la que se aplique (tabla 32.3).

## BIBLIOGRAFÍA DE LOS EJEMPLOS

Cannon CP, Braunwald E, McCabe CH, Rader DJ, Rouleau JL, Belder R, et al. Intensive versus moderate lipid lowering with statins after acute coronary syndromes. *N Engl J Med.* 2004;350:1495-504.  
Chan KB, Man-Son-Hing M, Molnar FJ, Laupacis A. How

well is the clinical importance of study results reported? An assessment of randomized controlled trials. *CMAJ.* 2001;165:1197-202.  
Duff JM, Leather H, Walden EO, LaPlant KD, George TJ. Adequacy of published oncology randomized

controlled trials to provide therapeutic details needed for clinical application. *J Natl Cancer Inst.* 2010;102(10):702-5.  
Duffy SW, Treasure FP. Potential surrogate endpoints in cancer research - some considerations and examples. *Pharmaceut Statis.* 2011;10:34-9.

- Hernández AV, Boersma E, Murray GD, Habbema JD, Steyerberg EW. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? *Am Heart J.* 2006;151:157-64.
- La Rosa JC, Grundy SM, Waters DD, Shear C, Barter P, Fruchart JC, et al. Intensive lipid lowering with atorvastatin in patients with stable coronary disease. *N Engl J Med.* 2005;352:1425-35.
- Psaty BM, Smith NL, Siscovick DS, Koepsell TD, Weiss NS, Heckbert SR, et al. Health outcomes associated with antihypertensive therapies used as first-line agents. A systematic review and meta-analysis. *JAMA.* 1997;277:739-45.
- Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet.* 1995;345:161-2.
- Steering Committee of the Physicians' Health Study Research Group. Preliminary report: findings form the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med.* 1988;318:262-4.
- Steg PG, López-Sendón J, López de Sa E, Goodman SG, Gore JM, Anderson FA, et al. External Validity of Clinical Trials in Acute Myocardial Infarction. *Arch Intern Med.* 2007;167(1):68-73.
- Travers J, Marsh S, Williams M, Weatherall M, Caldwell B, Shirtcliffe P, et al. External validity of randomised controlled trials in asthma: to whom do the results of the trials apply? *Thorax.* 2007;62:219-23.
- Van Spall HGC, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA.* 2007;297:1233-40.

## BIBLIOGRAFÍA

- Altman DG, Bland JM. Generalization and extrapolation. *BMJ.* 1998;317:409-10.
- Bailey KR. Generalizing the results of randomized clinical trials. *Control Clin Trials.* 1994; 15:15-23.
- Fayers PM, Hand DJ. Generalization from phase III clinical trials: survival, quality of life, and health economics. *Lancet.* 1997;350:1025-7.
- Fleming TR, De Mets DL. Surrogate end-points in clinical trials: are we being misled? *Ann Intern Med.* 1996;125:605.
- Fletcher J. Subgroup analyses: how to avoid being misled. *BMJ.* 2007;335:96-7.
- García Alonso F, Guallar E, Bakke OM, Carné X. El placebo en ensayos clínicos con medicamentos. *Med Clin (Barc).* 1997;109:797-801.
- Glasziou PP, Irwig LM. An evidence-based approach to individualising treatment. *BMJ.* 1995;311:1356-8.
- Guyatt GH, Juniper EJ, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. *BMJ.* 1998;316:690-3.
- Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? *JAMA.* 1994;271:59-63.
- Guyatt GH, Sinclair J, Cook DJ, Glasziou P, for the Evidence-Based Medicine Working Group and the Cochrane Applicability Methods Working Group. User's guides to the medical literature XVI. How to use a treatment recommendation. *JAMA.* 1999;281:1836-43.
- Haynes RB, Sackett DL, Gray JMA, Cook DJ, Guyatt GH. Transferring evidence from research into practice: 2. Getting the evidence straight. *ACP J Club.* 1997;126(1):A14-6.
- Jaeschke R, Guyatt GH, Sackett DL, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature III. How to use an article about a diagnostic test. B. What were the results and will they help me in caring for my patients? *JAMA.* 1994;271:703-7.
- Kent DM, Rothwell PM, Ioannidis JPA, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials.* 2010;11:85.
- Kottow M. The improper use of research placebos. *J Eval Clin Pract.* 2010;16:1041-4.
- Longford NT. Selection bias and treatment heterogeneity in clinical trials. *Stat Med.* 1999;18:1467-74.
- McQuay HJ, Moore RA. Using numerical results of systematic reviews in clinical practice. *Ann Intern Med.* 1997;126:712-20.
- Nieto García FJ, Peruga Urrea A. Riesgo atribuible: sus formas, usos e interpretación. *Gac Sanit.* 1990;4:112-7.
- Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet.* 2005;365:176-86.
- Rothwell PM. External validity of results of randomized trials: disentangling a complex concept. *Int J Epidemiol.* 2010;39:94-6.
- Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA.* 1995;273:408-12.

Scott IA, Greenberg PB.  
Cautionary tales in the clinical interpretation of therapeutic trial reports. *Int Med J.* 2005;35:611-21.

Simes RJ, Gebski VJ, Keech AC.  
Subgroup analysis: application to individual patient decisions. *Med J Aust.* 2004;180:467-9.

Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ.* 2010;340:c117.

Yudkin J, Lipska K, Montori V. The idolatry of the surrogate. *BMJ.* 2011;343:d7995.

# Parte

5

## Publicación

**33** Comunicación científica **311**

**34** El artículo original **316**

Página deliberadamente en blanco

## Comunicación científica

La publicación de artículos en las revistas biomédicas es el principal mecanismo que usan los investigadores para dar a conocer a la comunidad científica los resultados de los estudios que realizan y, por ello, es el medio más importante de difusión de los nuevos progresos de la medicina y de actualización de los conocimientos de los profesionales. Un estudio no se puede considerar finalizado hasta que no haya sido difundido en el ámbito que pueda ser de interés.

Para que la comunicación científica sea realmente eficaz, es necesario no sólo que los estudios se hayan efectuado con rigor metodológico, sino que también los artículos hayan sido redactados y publicados de forma correcta, para que permitan a los lectores evaluar tanto la validez de los resultados como su aplicabilidad a su propia práctica.

En el proceso de la comunicación científica escrita son cuatro los actores implicados: los investigadores o autores de los manuscritos, los editores de las revistas, los asesores que revisan los trabajos y los lectores de los artículos que se publican.

### AUTORES

Los investigadores son responsables de la veracidad y la originalidad de la información que presentan, y en caso necesario deben ser capaces de acreditarlas. No sólo deben haber diseñado y ejecutado los estudios con el rigor adecuado, sino que han de escribir los informes con honestidad, sin manipular la información, inventar ni omitir datos para dar la sensación de que su trabajo es mejor de lo que realmente es.

La Declaración de Helsinki establece que los autores tienen el deber de poner a la disposición del

público los resultados de su investigación en seres humanos y que son responsables de la integridad y exactitud de sus informes, y recuerda que se deben publicar tanto los resultados negativos e inconclusos como los positivos o, en caso contrario, deben estar a la disposición del público.

Los investigadores están sometidos a múltiples presiones para publicar, ya sea por metas personales o exigencias institucionales, por ejemplo, lo que conduce en ocasiones a conductas inapropiadas, como hacer constar como autores a personas que no lo son (autoría regalada) o a personas inexistentes (autoría «fantasma»), o no hacer constar a todos los verdaderos autores del trabajo.

Según el Comité Internacional de Editores de Revistas Médicas ([www.icmje.org](http://www.icmje.org)), cada autor debería haber participado suficientemente como para tomar la responsabilidad pública sobre partes del contenido del artículo. La autoría debe basarse sólo en contribuciones sustanciales a:

1. La concepción y el diseño o el análisis y la interpretación de los resultados.
2. La redacción y revisión crítica del contenido intelectual del artículo.
3. La aprobación de la versión final.

Concreta, además, que la adquisición de fondos, la recogida de datos o la supervisión general del equipo investigador por sí solos no justifican la autoría.

Debido a que cada vez los trabajos son más complejos y requieren la participación de equipos numerosos y multidisciplinarios de investigadores, resulta difícil que existan personas que cumplan estrictamente con todos los requisitos citados. Por ello, existe una tendencia a solicitar que los firmantes de un trabajo especifiquen cuál ha sido su contribución

al estudio (diseño, recogida de datos, análisis de los resultados, obtención de fondos, revisión intelectual del manuscrito, etc.).

Cuando los móviles para publicar son egoístas, como el afán de aumentar el currículum o el prestigio personal, conseguir una promoción profesional o mantener la rivalidad de grupos de trabajo, el investigador se expone al riesgo de realizar estudios poco originales y publicarlos de forma rutinaria. Es prácticamente inevitable que sigan viendo la luz artículos sin originalidad ni interés, ya que con frecuencia se equipara mayor mérito científico con mayor número de publicaciones, y se valora la cantidad por encima de la calidad.

Debe evitarse la publicación fragmentada, que consiste en descomponer un trabajo y publicar sus fragmentos de forma separada como artículos originales y la publicación duplicada o repetida, que ocurre cuando un artículo presenta esencialmente la misma información o con escasas variaciones que otro ya publicado. Es obligación del autor informar al editor de cualquier situación que pudiera ser motivo de publicación redundante. La publicación previa de un resumen o la presentación total o parcial de datos en una reunión científica no se consideran objeto de publicación redundante, aunque se aconseja que este hecho conste al pie de la primera página del manuscrito.

El Comité Internacional de Editores de Revistas Médicas ([www.icmje.org](http://www.icmje.org)) sólo considera aceptable la publicación secundaria, especialmente en otro idioma, cuando se cumplen las siguientes condiciones:

1. Los autores tienen la aceptación de los editores de ambas revistas.
2. Se respeta la prioridad de la publicación primaria.
3. El artículo se dirige a un grupo diferente de lectores, por lo que una versión abreviada suele ser suficiente.
4. La versión secundaria refleja fielmente los datos y la interpretación de la publicación primaria.
5. Una nota al pie de página informa de la publicación previa y la referencia.

También debe evitarse la inflación de autoría, que algunos denominan autoría «fantasma», que consiste en hacer constar como coautores a personas que no han contribuido de forma sustancial al estudio. Es importante respetar escrupulosamente los criterios de autoría.

Los autores deberían hacer constar la existencia de cualquier relación con empresas o instituciones u otro tipo de situación que pudiera suponer un conflicto de intereses y comprometer la interpretación objetiva de los datos. Además, deben redactar los artículos con claridad, pensando en los lectores

potenciales y respetando las recomendaciones internacionales para la presentación de manuscritos a las revistas biomédicas.

## EDITORES

Los editores deben determinar la línea editorial de la revista. Su función, y la del consejo de redacción, no consiste en poner dificultades a los autores, sino en ayudarles a publicar artículos de contenido y estilo correctos. Por ello, es recomendable que las revistas publiquen periódicamente las instrucciones para los autores, así como información sobre el funcionamiento del proceso editorial y de evaluación de los trabajos.

La expansión que ha experimentado el volumen de revistas periódicas para dar cabida al ingente número de publicaciones ha supuesto un enorme esfuerzo para mantener la calidad, preservar la validez científica del texto impreso y proteger el derecho del lector a ser correctamente informado. En la actualidad se considera clave para garantizar la calidad de una revista que los trabajos que se reciben sean sometidos a un proceso de evaluación por expertos que no formen parte del consejo editorial (*peer-review*). Además, los editores han de acoger de buen grado la publicación de los errores que han de ser corregidos, de notificaciones de los autores, y de cartas de lectores que llaman la atención sobre el contenido erróneo o discutible de lo publicado. En caso necesario, se publicarán con el debido relieve las retracciones acerca de artículos falsificados o plagiadoss.

## REVISORES

Los objetivos principales del proceso de revisión por expertos (*peer-review*) son evitar la publicación de un trabajo de mala calidad científica, no original o que no contenga información relevante para los lectores de la revista, así como mejorar la redacción y la presentación de los datos del artículo.

El proceso de revisión ha recibido diversas críticas, entre las que destacan que es costoso, que consume mucho tiempo a los expertos, que éstos pueden tener un sesgo en contra o a favor del mensaje del artículo, o incluso un conflicto de intereses. El revisor puede no tener la competencia técnica o los conocimientos metodológicos suficientes para valorar la calidad del diseño de un estudio y/o la relevancia de la información que se presenta. Además, el proceso de revisión puede introducir un retraso innecesario en la comunicación de hallazgos importantes, a veces incrementado por la pereza del revisor que incumple los plazos de tiempo que marcan los directores de las revistas.

El proceso de revisión científica se basa en la confianza y en la responsabilidad. Los revisores son responsables de evaluar el contenido científico del trabajo, sus méritos o defectos intrínsecos de forma imparcial, desde una postura de máxima neutralidad, y de acuerdo con los estándares de calidad. El evaluador no puede cometer abusos de poder, tomando decisiones rápidas y caprichosas, ni menospreciar u ofender a los autores. Las críticas al estudio deben realizarse siempre con corrección y profesionalidad, respetando el pluralismo científico y sin valoraciones peyorativas de los investigadores. Los asesores son colaboradores tanto del director como del autor, jamás sus adversarios.

La competencia científica y técnica es una condición indispensable para ser un buen evaluador. La obligación de competencia incluye el deber de estudiar a fondo el artículo con el fin de dar un dictamen en conciencia. Si el experto no se considera competente para evaluar el trabajo, deberá notificarlo, sin pérdida de tiempo, al director de la revista.

La documentación que se entrega a los evaluadores para su revisión es confidencial. El artículo todavía inédito no puede ser divulgado por el experto. No puede hacer copias de él ni pasarlo a sus colaboradores u otras personas interesadas sin las autorizaciones y garantías adecuadas.

## LECTORES

La finalidad de la comunicación científica en medicina no es simplemente la publicación de un estudio de investigación u otro tipo de artículo, sino la mejora de la atención sanitaria mediante la aplicación en la práctica clínica de los resultados de dichos estudios y de las recomendaciones que de ellos se derivan.

**Ejemplo 33.1.** Es muy ilustrativo el trabajo realizado por Antman et al (1992) sobre el tratamiento trombolítico en el infarto agudo de miocardio. Estos autores comprobaron que, si bien la evidencia disponible en la literatura científica entre 1970 y 1975 ya era suficiente para demostrar que la intervención era efectiva, las revisiones y los libros de texto publicados en esas fechas no mencionaban este tratamiento, o a lo sumo lo presentaban como experimental. No fue hasta 1980 que algún texto realizó una recomendación específica, y sólo posteriormente a 1985 se empezó a recomendar su uso habitual, es decir, prácticamente 15 años después de que la evidencia publicada fuera suficiente.

El lector debe ser consciente de que no todos los resultados que se publican son válidos, que no

todos los resultados válidos son relevantes, y que no todos los resultados válidos y relevantes son aplicables a su propia práctica clínica. Por ello, debe ser capaz de leer críticamente la literatura y evaluar tanto la validez como la relevancia y la aplicabilidad de los artículos.

Dado el enorme volumen de artículos que se publican, han aparecido algunas revistas que tienen por finalidad ayudar al lector en su tarea de estar al día, de modo que seleccionan los artículos que presentan resultados válidos y de mayor relevancia, los resumen en un formato uniforme y elaboran un comentario crítico de los aspectos relacionados con su aplicación práctica.

El resultado de la investigación es conocimiento, pero para que pase a formar parte de la ciencia, este conocimiento individual debe presentarse de manera que su validez pueda juzgarse de forma independiente. No puede considerarse que una investigación ha finalizado hasta que no se ha difundido a la comunidad científica, lo que habitualmente significa su publicación en una revista y su inclusión en las bases de datos bibliográficas.

En todas las fases de la comunicación científica son múltiples los conflictos de intereses que pueden presentarse. Dado que la literatura científica es la principal fuente de evidencias que se utiliza en la toma de decisiones y en la elaboración de las guías de práctica clínica y las recomendaciones para la atención de los pacientes, el impacto que estos conflictos pueden tener sobre la práctica es muy importante.

El Committee On Publication Ethics (COPE) ha elaborado unas guías sobre las buenas prácticas en publicación ([www.publicationethics.org.uk](http://www.publicationethics.org.uk)) que abordan diez aspectos:

1. Diseño del estudio y aprobación ética.
2. Análisis de los datos.
3. Autoría.
4. Conflicto de intereses.
5. Revisión de manuscritos.
6. Publicación redundante.
7. Plagio.
8. Deberes de los editores.
9. Relaciones con los medios de comunicación.
10. Publicidad.

El COPE define conflicto de interés como aquel que, si se revela posteriormente, podría hacer que el lector se sintiera razonablemente decepcionado o engañado, incluyendo también los conflictos no aparentes que pueden influir sobre el juicio del autor, los revisores o los editores.

Los conflictos pueden manifestarse de múltiples formas: falsificación o fabricación de datos, publicación

selectiva de información, decisión de publicar o no un trabajo, interpretación sesgada de los resultados, etc.

**Ejemplo 3.2.** Stelfox et al (1998) identificaron los artículos publicados entre marzo de 1995 y septiembre de 1996 relacionados con la seguridad de los fármacos calcioantagonistas, los clasificaron como favorables, neutrales o críticos en relación con su uso, y solicitaron información de los autores de dichos trabajos sobre sus relaciones con la industria farmacéutica. El principal resultado fue que los autores favorables al uso de estos fármacos tenían relaciones financieras con la industria farmacéutica en un tanto por ciento superior al de los que eran neutrales o críticos (el 100% frente al 67 y el 43%, respectivamente).

**Ejemplo 3.3.** En una revisión sistemática reciente, Lexchin et al (2003) demostraron la existencia de un sesgo, de manera que los estudios financiados por

la industria farmacéutica tienen una mayor probabilidad de obtener resultados favorables para el fármaco que producen. Este sesgo no está relacionado con diferencias en la calidad metodológica de los estudios, sino que las explicaciones más probables son la elección de un comparador inadecuado y el sesgo de publicación.

Este problema ha llegado a ser tan importante que los editores de las principales revistas médicas publicaron conjuntamente un artículo editorial (Davidoff et al, 2001) en el que abogaban por la declaración por parte de todos los implicados en el proceso de revisión de manuscritos y de publicación de cualquier relación que pudiera ser vista como un potencial conflicto de interés. Insisten en que ningún investigador debería participar en acuerdos que interfieran en su accesibilidad a los datos o a su capacidad para analizarlos independientemente, preparar manuscritos y publicarlos.

## BIBLIOGRAFÍA DE LOS EJEMPLOS

Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analysis of randomized control trials and recommendations of clinical experts. *Treatments for*

myocardial infarction. *JAMA*. 1992;268:240-8.  
Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ*. 2003;326:1167-76.

Stelfox HT, Chua G, O'Rourke K, Detsky AS. Conflict of interest in the debate over calcium-channel antagonists. *N Engl J Med*. 1998;338:101-6.

## BIBLIOGRAFÍA

Argimon JM, Jiménez J, Martín Zurro A, Vilardell M. La literatura científica biomédica. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 1-13.  
Cardellach F, Ribera JM. El proceso editorial. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 309-26.

Cobo E, Cortés J, Ribera JM, Cardellach F, Selva-O'Callaghan A, Kostov B, et al. Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: masked randomised trial. *BMJ*. 2011;343:d6783.  
Davidoff F, DeAngelis CD, Drazen JM, Hoey J, Høgaard L, Horton R, et al. Sponsorship, authorship, and accountability. *JAMA*. 2001;286:1232-4.  
Gisbert JP, Piqué JM. Autoría. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo

escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 327-45.  
*JAMA and Archives Journals. AMA manual of style* 10th ed. New York: Oxford University Press; 2007.  
Lorenzo S, Carrasco G. El sistema de revisión por expertos (peer review) en las revistas científicas: ventajas y limitaciones. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 291-307.

Pascual A, Martínez Martínez L. Mala conducta y conflicto de intereses. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup>

ed. Barcelona: Elsevier; 2010. p. 347-61.

Simera I, Altman DG, Moher D, Schulz KF, Hoey J. La red EQUATOR: facilitando la comunicación transparente y precisa de la investigación sanitaria. En: Jiménez Villa

J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 265-71.

## El artículo original

Los artículos médicos están sujetos a reglas estrictas de redacción y publicación y deben seguir las recomendaciones del Comité Internacional de Editores de Revistas Médicas, conocidas popularmente como normas de Vancouver, porque fue en este lugar donde un grupo de editores de algunas de las principales revistas biomédicas se reunieron para uniformar los requisitos técnicos de los manuscritos presentados para su publicación. Su finalidad principal es conseguir que la comunicación entre el investigador y el lector del trabajo sea coherente, clara y precisa.

Existen diversas formas de comunicación científica, como la oral o la escrita, y diferentes soportes, como el papel o el medio electrónico. No es lo mismo presentar una comunicación oral en un congreso que un póster, o que participar en una mesa redonda. Tampoco se escribe igual un editorial, un artículo original, una revisión, una nota clínica o una carta al director. De la misma forma, el enfoque debe ser distinto si el texto va destinado a una monografía, una revista de medicina general, una publicación especializada o si se va a difundir por internet.

Sin embargo, una norma común a todos ellos es que los autores deben proporcionar la suficiente información para que los lectores puedan evaluar los resultados y las conclusiones del estudio, o repetirlo en el supuesto de que otros investigadores estén interesados en el tema y lo consideren oportuno.

En el contexto de este libro sería demasiado extenso analizar detalladamente las reglas que rigen todas estas formas de expresión oral y escrita. Este capítulo se limita a presentar las normas para la elaboración de un artículo original, ya que es el tipo de difusión más habitual de una investigación clínica y epidemiológica.

Un artículo original es un trabajo publicado por primera vez por su autor, a partir de la investigación llevada a cabo en un tema determinado. Su estructura habitual se conoce mediante el acrónimo IMRD, que representa las iniciales de sus apartados: Introducción, Material y métodos, Resultados y Discusión ([tabla 34.1](#)). Además, consta del título, el resumen y las referencias bibliográficas.

Antes de redactar el artículo, conviene decidir la revista a la que se enviará el manuscrito, con el fin de que la redacción sea comprensible y atractiva para sus lectores, y poder consultar sus instrucciones para autores y requisitos específicos.

### INTRODUCCIÓN

Su propósito general es centrar el tema. Para ello, el autor debe indicar claramente las bases en que se fundamenta el estudio, la laguna de conocimientos que pretende llenar o, en otras palabras, las razones que justifican su realización y su objetivo específico.

La introducción debe redactarse pensando en realizar la necesidad del estudio y estimular la curiosidad del lector. Una característica importante es la brevedad en la exposición y se considera que, por regla general, esta sección no debe superar una página y media (45 líneas). Debe centrar el tema, no revisarlo. Hay que utilizar pocas citas bibliográficas, y éstas deben ser de actualidad y estar bien seleccionadas. Es recomendable que el número de referencias bibliográficas de este apartado no supere la mitad del total.

Hay que evitar realizar una revisión histórica de la enfermedad y explicar lo que se puede encontrar en cualquier texto de consulta. Si un lector ha decidido

**Tabla 34.1** Formato IMRD de un artículo original

<b>I</b>	Introducción	¿Qué se ha estudiado?
<b>M</b>	Material y métodos	¿Cómo se ha estudiado?
<b>R</b>	Resultados	¿Qué se ha observado?
<b>D</b>	Discusión	¿Qué interpretación tienen los datos?

leer este artículo, es porque está interesado en el problema y ya conoce sus aspectos más generales. Por esta razón no debe proporcionársele información genérica, sino la específica del problema concreto que se ha estudiado. Si un lector quiere tener una visión general de una enfermedad, escogerá artículos de revisión o una monografía. Por este mismo motivo, se recomienda no citar libros de texto en la bibliografía.

Cuando se concibe una pregunta y se planifica una investigación, debe asegurarse que la pregunta que se formula no ha sido contestada previamente o que lo ha sido de forma contradictoria. Es frecuente que un investigador novel se desilusione cuando, después de concebir una idea, busque en la literatura y encuentre que ya se ha investigado antes sobre el tema. Siempre hay que pensar qué aspectos novedosos aportará la investigación. Incluso cuando la idea ya ha sido previamente investigada, es útil preguntarse si el método empleado por los anteriores investigadores era el más adecuado, o bien si puede considerarse útil repetirlo en otras poblaciones. Las razones que han llevado a los investigadores a realizar el estudio deben ser comentadas de forma concisa en la introducción, sin abrumar a los lectores con multitud de datos irrelevantes sobre la trascendencia del problema. Tampoco deben adelantarse acontecimientos, presentando, en este apartado, datos y conclusiones que se deriven del trabajo realizado.

El último párrafo de la introducción se destina a enunciar el objetivo del estudio. Su definición explícita y clara facilita la comprensión del artículo, ya que el lector puede identificar con facilidad sus aspectos más relevantes.

En el cuadro 34.1 se resumen los errores más frecuentes que aparecen en la introducción de un artículo.

## MATERIAL Y MÉTODOS

Su finalidad es describir el diseño de la investigación y cómo se ha llevado a cabo. Es el apartado evaluado con mayor detalle en el proceso de revisión por los

**Cuadro 34.1** Errores frecuentes en la introducción de un artículo original

- Realizar una revisión exhaustiva del tema
- Excesivo número de referencias bibliográficas
- Poca actualidad de las referencias bibliográficas
- Iniciar la introducción definiendo la enfermedad en estudio
- Realizar una revisión histórica de la enfermedad
- Explicar lo que se puede encontrar en cualquier texto de consulta
- Ausencia de justificación del estudio
- Abrumar al lector con multitud de datos irrelevantes sobre la trascendencia del problema
- Omitir el objetivo o enunciarlo de forma inadecuada

asesores de la revista y el más vulnerable para el rechazo de un manuscrito. Aunque habitualmente se denomina *Material y métodos*, algunos autores consideran que no es una expresión adecuada cuando la investigación se ha llevado a cabo en seres humanos y sugieren denominarlo *Pacientes y métodos*, o bien *Población y métodos* si se han estudiado sujetos presuntamente sanos.

Probablemente, el principal problema que se plantean los autores al redactar este apartado es decidir cuánta información debe incluirse y con qué detalle debe presentarse. La respuesta es que debe incluirse la información suficiente para que un lector pueda comprender adecuadamente todo el proceso de la investigación, y con el detalle suficiente para que otros investigadores puedan replicar el estudio. Por lo tanto, el texto de este apartado debe incluir los párrafos correspondientes al diseño, los sujetos, las fuentes de información, las intervenciones, los instrumentos, las variables de respuesta, los procedimientos y la estrategia de análisis estadístico. Para facilitar la lectura, especialmente si es extenso, pueden utilizarse subpáginas para cada uno de los bloques de información citados.

El primer párrafo se reserva a la descripción, en una o dos frases, del tipo de estudio realizado. A continuación se especifican y describen los criterios de inclusión y exclusión seguidos para seleccionar a la población de estudio. La definición de la enfermedad o de la exposición que se estudia es el criterio que debe definirse con mayor precisión. Las finalidades de esta descripción son evaluar la adecuación de la población, conocer en qué tipo de individuos se ha observado la presencia o la ausencia del efecto o de la asociación estudiados, evaluar la posibilidad de extrapolar los resultados a otras poblaciones y permitir la replicabilidad del estudio.

También debe describirse la procedencia de los sujetos, ya que para la interpretación de los resultados y su extrapolación no es lo mismo que procedan de un centro de atención primaria o de un hospital, por ejemplo, así como la técnica utilizada para la captación de sujetos, indicando aquellos detalles que permitan conocer las características de la muestra de sujetos incluida (si son pacientes derivados, si se captan directamente de la consulta o a partir de algún sistema de registro, si existen filtros previos a su inclusión, etc.). En caso de haberse utilizado alguna técnica de muestreo, debe describirse adecuadamente. Es conveniente especificar el tiempo empleado en reunir a todos los participantes.

Hay que indicar qué asunciones se han utilizado para calcular el número de sujetos necesarios para la realización del estudio.

En los estudios experimentales hay que especificar el método de asignación de los participantes a los grupos de estudio.

Una vez definidos los aspectos relacionados con la población de estudio y la formación de los grupos, se describen las variables estudiadas, incluyendo su definición, las técnicas de medida y fuentes de información utilizadas, y si fuera necesario, estimaciones o comentarios acerca de su validez y precisión.

También debe describirse la arquitectura general del estudio, indicando la pauta de visitas y qué actividades se realizan en cada una de ellas, quiénes y cuántos son los observadores, si ha existido un entrenamiento previo, los mecanismos de recaptación de los pacientes, las técnicas de control de calidad de los datos recogidos, etc. Si el estudio evalúa la eficacia de un tratamiento, debe señalarse si ha existido un período de lavado o blanqueo previo a la intervención, las técnicas de registro de la presencia de reacciones adversas, los criterios utilizados para la retirada anticipada de sujetos, las técnicas de enmascaramiento utilizadas (simple ciego, doble ciego, triple ciego, etc.). Además, también debe especificarse si el estudio ha sido aprobado por el Comité Ético de Investigación Clínica correspondiente.

La última sección se reserva a la descripción de la estrategia de análisis y las técnicas estadísticas empleadas. Cuando se utilizan varias pruebas ha de quedar claro en qué parte del artículo se utiliza cada una de ellas. Si las pruebas son de uso muy frecuente, no es necesario que se describan en el texto. Si su uso es más infrecuente, conviene incluir una breve explicación acompañada de una cita bibliográfica. Es preferible que las referencias bibliográficas correspondan a un libro de texto o a un artículo de revisión sobre la prueba, antes que recurrir al artículo original. Este último tiene, indudablemente, un gran valor desde el punto de

vista histórico y metodológico, pero suele ser muy complicado y ofrece pocos ejemplos prácticos. Por el contrario, los libros de texto o los artículos de revisión suelen ser más didácticos, lo cual facilita la comprensión del lector. Es recomendable hacer constar el programa estadístico utilizado. Hay que indicar la naturaleza de la hipótesis evaluada (unilateral o bilateral) y el valor de significación aceptado (habitualmente el 0,05).

Dado que se está relatando lo que sucedió durante la investigación, lo más adecuado es redactar este apartado en tiempo pasado. El estilo debe ser directo, utilizando frases cortas y sencillas que describan el proceso de la investigación con claridad, concisión y en la secuencia lógica. En el cuadro 34.2

#### Cuadro 34.2 Principales errores cometidos en la redacción del apartado de Material y métodos

- No identificación del diseño del estudio
- Identificación de un diseño que no se corresponde con el realmente utilizado
- Diseño inapropiado para los objetivos de la investigación
- Descripciones innecesarias o superfluas
- Ausencia de descripción del universo o población de referencia
- Ausencia de descripción del proceso de captación de los sujetos
- Definición insuficiente de los criterios de selección de los sujetos
- Ausencia de información sobre el cálculo del tamaño de la muestra o sobre las asunciones en que se ha basado
- Ausencia de datos sobre la asignación de los sujetos a los grupos de estudio
- Definición de variables que posteriormente no se utilizan en el análisis (frecuente en artículos que informan sobre aspectos parciales de una investigación más amplia)
- Explicación detallada de variables o pruebas complementarias no esenciales dentro del estudio
- Omisión de información necesaria para la interpretación del estudio o utilizar descripciones incompletas (p. ej., criterios de selección, definición de las principales variables del estudio, descripción de la intervención, etc.)
- Inclusión anticipada e innecesaria de resultados
- Omisión de las pruebas estadísticas utilizadas
- Especificación de la marca y el tipo de ordenador utilizado
- Ausencia de orden lógico en la descripción de los procedimientos

se presentan los principales errores que se cometen al redactar este apartado.

## RESULTADOS

Su finalidad es presentar los hallazgos que los autores han encontrado en su investigación y los resultados de su análisis, pero no interpretarlos.

Esta sección debe contener sólo la información importante, pero sin omitir nada que pueda interesar al lector o que sea necesario para la adecuada comprensión de los hallazgos. A algunos autores les cuesta mucho decidir qué información merece la pena incluir y, ante la duda, facilitan todo lo que ha estado a su alcance. Publicar toda la información disponible no es señal de honestidad científica, sino de falta de criterio en su selección.

Es conveniente evitar la exposición excesivamente detallada y exhaustiva de los datos numéricos, y recurrir a la síntesis, organización y presentación en tablas y figuras, cuando sea adecuado, teniendo siempre en cuenta que su función es complementar el texto, no duplicarlo. En el texto deben resaltarse los hallazgos principales, es decir, aquellos que se refieren a las hipótesis de trabajo y en los que se basarán las conclusiones, y remitir al lector a las tablas y figuras en las que encontrará organizado el resto de los datos relevantes.

Los resultados se han de presentar siguiendo una sucesión lógica, de acuerdo con la estrategia de análisis que se ha empleado. En primer lugar se expondrán los datos descriptivos de las principales características de los sujetos estudiados. En los estudios analíticos, a continuación, se evalúa la comparabilidad inicial de los grupos de estudio respecto a las variables que pueden influir sobre la respuesta. Esta evaluación no debe basarse exclusivamente en criterios de significación estadística, ya que no resulta infrecuente observar diferencias estadísticamente significativas pero clínicamente irrelevantes, especialmente si los grupos son de gran tamaño, o bien diferencias importantes que pueden condicionar las respuestas, pero que no resultan estadísticamente significativas debido a que el número de sujetos es reducido. Por ello, en la evaluación de la comparabilidad debe prestarse mucha atención a la magnitud de las diferencias existentes entre los grupos y a su posible impacto sobre la respuesta.

A continuación se presenta el resultado principal, que corresponde a la estimación del efecto del factor de estudio sobre la variable de respuesta, o bien a la estimación de la asociación entre ambas variables. La elección de la medida adecuada para presentar este resultado es importante, ya que los

lectores pueden verse influidos por la forma de presentación utilizada. Así, un resultado expresado en forma de reducción del riesgo relativo suele tener una mayor influencia sobre las decisiones de los profesionales que el mismo resultado expresado en forma de reducción absoluta de riesgo o de número de personas que es necesario tratar. En caso de que existan potenciales factores de confusión, debe presentarse la estimación del efecto o de la asociación ajustada por ellos.

A continuación, se presentan los resultados del análisis de subgrupos, si se ha realizado. Y, finalmente, los relacionados con las preguntas secundarias del estudio.

En cualquier estudio hay que informar del número de no respuestas o de personas perdidas durante el seguimiento, así como de los motivos por los que se han producido.

Debe prestarse especial atención al uso inadecuado de determinados términos, tales como «normal», «aleatorio», «significativo» o «correlación», con un significado estadístico muy concreto, pero que son utilizados inadecuadamente con un significado menos técnico.

La información sobre los resultados de las pruebas estadísticas debe incluir la prueba utilizada, el valor del criterio estadístico, los grados de libertad y el valor de significación estadística, así como los intervalos de confianza.

En los cuadros 34.3 y 34.4 se presentan los errores más frecuentes en la presentación de los datos estadísticos y en el apartado de Resultados, respectivamente.

### Tablas

Las tablas son conjuntos ordenados y sistemáticos de números y/o palabras presentados en filas y columnas. Su finalidad es la ordenación y presentación de información de tipo repetitivo de manera comprensible. Constituyen un complemento muy útil del texto, pero no deben duplicarlo.

Las tablas deben ir numeradas de forma consecutiva, según el orden en que se mencionan en el texto. Deben presentarse separadamente del texto del artículo, cada una en una hoja aparte. Así mismo, deben tener un título breve (habitualmente menos de diez palabras) que describa su contenido, sin adelantar información sobre los resultados, y que, salvo casos especiales, no debe incluir abreviaturas.

La facilidad y rapidez de comprensión de una tabla depende en gran medida de la lógica de su construcción (cuadro 34.5). Las filas y las columnas deben ir precedidas de un encabezamiento corto o abreviado, que identifique exactamente el material

### Cuadro 34.3 Errores frecuentes en la presentación de resultados estadísticos

- $p$  huérfana (presentar el grado de significación estadística sin indicar la prueba utilizada)
- Presentar estimaciones puntuales sin intervalos de confianza
- Expresar el grado de significación estadística de forma dicotómica: significativo o no significativo. Es preferible indicar el valor de  $p$ , especialmente cuando no está muy alejado del valor 0,05
- Presentar solamente el grado de significación estadística, sin una estimación de la magnitud de la diferencia, efecto o asociación
- Precisión excesiva de los resultados (p. ej.,  $p=0,000000002$ , o edad media = 34,6785 años)
- Expressar los resultados en porcentajes cuando el número de casos es muy reducido (p. ej., en un estudio sobre 8 pacientes, escribir: el 50% de los casos...)
- Acompañar una media aritmética de  $\pm$  sin indicar si corresponde a la desviación estándar, error estándar o intervalo de confianza
- Utilizar el error estándar en lugar de la desviación estándar cuando se describe la distribución de una variable

### Cuadro 34.4 Errores frecuentes en el apartado de Resultados

- Incluir información no pertinente (no relacionada con los objetivos del estudio)
- Incluir información irrelevante
- Omitir información relevante
- Presentar los resultados sin una secuencia lógica
- Interpretar los resultados que se presentan o acompañarlos de opiniones de los autores
- Presentación inadecuada de los aspectos estadísticos (ver cuadro 34.3)
- Duplicar la información presentada en tablas, figuras o texto
- No citar todas las tablas y figuras que presentan resultados

que contienen. El uso de símbolos y/o abreviaturas puede suponer un deseable ahorro de espacio y una simplificación de la presentación. En caso de utilizarse, debe recurrirse a las abreviaciones estándar y los símbolos convencionales, cuyo significado, si no es autoexplicativo, debe hacerse explícito mediante el uso de notas a pie de tabla.

### Cuadro 34.5 Recomendaciones generales para la elaboración de tablas

- No deben elaborarse a menos que deban presentarse datos repetitivos
- No deben duplicar el texto. Los datos se presentan en texto, en tablas o en figuras, pero no en más de una forma a la vez
- Han de ser autoexplicativas
- Han de ser sencillas y de fácil comprensión
- El título debe ser breve y claro
- Los encabezamientos de las filas y columnas deben ser claros e incluir las unidades de medida
- La ordenación de los datos debe ser de izquierda a derecha, y de arriba abajo. Las filas y columnas deben presentar una ordenación lógica
- Debe evitarse la utilización de un número excesivo de decimales
- Deben incluirse los valores marginales (totales) para facilitar la comprensión
- Si se utilizan abreviaturas, deben hacerse explícitas
- Utilizar una disposición similar en tablas similares

Una buena tabla debe ser autoexplicativa, en el sentido de que la información que presenta sea autosuficiente para su comprensión, incluso separadamente del texto.

Cuando se presenta más de una tabla de resultados, es conveniente utilizar un formato similar en todas ellas, de forma que se facilite su comprensión por el lector. Por otro lado, debe comprobarse que todas aparecen mencionadas en el texto, y que los datos que se presentan concuerden con los que se citan.

### Figuras

Una figura es un material de ilustración que incluye tanto gráficos que presentan datos numéricos en forma visual como diagramas o fotografías.

En muchas ocasiones, los autores dudan acerca de si es más adecuado presentar los datos en forma de tabla o de figura. Como norma general, en un artículo es preferible el uso de tablas, ya que son más exactas y permiten presentar mucha más información que un gráfico, a no ser que se deseé evidenciar la existencia de una tendencia en los datos. Lo que no es correcto es duplicar la información, presentando los mismos datos en ambas formas.

Las figuras se numeran consecutivamente según el orden en que se mencionan en el texto y deben

presentarse separadamente del mismo. Al igual que las tablas, deben tener un título breve y explicativo, que no debe incluirse en la figura, sino en una hoja aparte en la que consten los títulos, leyendas y notas que deben acompañar a las figuras del manuscrito.

Si se incluyen fotografías de personas, éstas no deben ser identificables, o, si lo son, deben ir acompañadas del correspondiente permiso escrito para su utilización.

La finalidad de la utilización de figuras en la presentación de datos cuantitativos es ayudar a su clara descripción, exploración, tabulación y comparación. Sin embargo, en su elaboración debe tenerse un especial cuidado en no distorsionar lo que se pretende mostrar, ya que la presentación visual puede inducir al lector a una interpretación errónea.

## DISCUSIÓN

En este apartado se interpretan los resultados, destacando los aspectos más novedosos y relevantes, analizándolos de forma cautelosa, relacionándolos con las teorías subyacentes, comparándolos de forma objetiva con los hallazgos de otros estudios, y argumentando las conclusiones que de todo ello se derivan. También debe evaluarse la validez interna del propio estudio, comentando sus limitaciones, así como la validez externa, es decir, el grado en que se pueden extraer o generalizar los resultados, exponiendo sus implicaciones teóricas y prácticas. Todo ello debe hacerse de forma honesta y objetiva, manteniendo un equilibrio entre la sobrevaloración y la subestimación de la importancia de los propios hallazgos.

Es útil comenzar con una recapitulación muy concisa de los hallazgos principales que constituyen la respuesta a la pregunta principal del estudio, para captar de este modo la atención del lector y evitar así que se convierta en una simple, inútil y pesada repetición de los resultados.

Deben comentarse honestamente las limitaciones del estudio. Nadie conoce mejor sus posibles fuentes de error que quien lo ha realizado. Por lo tanto, los autores deben hacer una autocrítica honesta de su trabajo, señalando sus puntos fuertes y débiles, realizando, cuando sea posible, recomendaciones para mejorar el método en estudios posteriores.

Los resultados observados deben compararse con los de otros trabajos que han utilizado un diseño y una metodología similares. Los estudios citados no se han de seleccionar previamente en función de si sus resultados son o no congruentes con los observados en el propio trabajo. Por otro lado, la alusión a otras publicaciones debe hacerse con precisión, sin

mutilar frases ni palabras, y exponiendo lo que el autor referenciado ha querido decir realmente. Deben explicarse claramente, si existen, las diferencias entre el trabajo original y el de otros autores, tanto en el método seguido como en los resultados. No deben ocultarse las diferencias con los resultados obtenidos en otros trabajos, sino que hay que tratar de explicarlas y, si no es posible, admitir la incapacidad para hacerlo.

Los autores han de discutir las implicaciones teóricas y prácticas del estudio con prudencia e imparcialidad. La discusión debe basarse en la estimación de la magnitud del efecto y su relevancia clínica, y no simplemente en el grado de significación estadística.

Finalmente, deben indicarse las líneas futuras de investigación e incluir todas aquellas deducciones que puedan sustentar un futuro estudio, evitando frases como el manido «es necesario efectuar nuevas investigaciones en este campo», sin especificar los aspectos concretos que convendría seguir trabajando.

En el cuadro 34.6 se presentan los errores más frecuentes que se cometen en la redacción de este apartado.

## BIBLIOGRAFÍA

Cuando se selecciona y escribe la bibliografía hay que tener en cuenta dos normas. La primera, no citar todos los artículos indiscriminadamente por el mero hecho de haber tenido acceso a ellos, sino escoger los que se consideren más adecuados para que los lectores puedan evaluar los resultados. La segunda es citar sin error.

Muchos autores piensan que un artículo adornado con decenas de citas bibliográficas posee más rigor científico. Además, creen que los trabajos originales deben ser artículos de revisión. Nada más alejado de la realidad. La tarea del autor es seleccionar, entre los centenares de artículos que ha reunido y revisado, las citas más relevantes, que conjuguen la originalidad con la validez metodológica. Además, es más fácil publicar estudios poco voluminosos. A igualdad de interés, los editores prefieren artículos cortos en los que ha existido un trabajo previo por parte de los autores de seleccionar las mejores y más novedosas referencias bibliográficas.

Las normas de Vancouver especifican cómo debe citarse cualquier tipo de documento, como artículos de revistas, libros, prensa o documentos en soporte electrónico, entre otros. Para evitar los errores en la citación, hay que comprobar siempre las referencias con el original delante para no omitir ninguna parte de la cita. Los errores en la transcripción de una referencia dificultan o impiden su localización por los

#### Cuadro 34.6 **Errores frecuentes en la elaboración de la Discusión de un artículo original**

- Realizar una discusión general, no centrada en los resultados de la investigación
- Convertir la discusión en una revisión del tema
- Repetir detallada e innecesariamente los resultados sin interpretarlos
- Repetir conceptos ya planteados en la introducción
- Exagerar la importancia de los resultados encontrados
- Conceder importancia solamente a la significación estadística, sin evaluar la magnitud del efecto observado
- No confrontar los resultados con los de otros estudios
- No relacionar los resultados con las hipótesis y objetivos del trabajo
- Realizar interpretaciones que no son congruentes con los resultados
- Realizar interpretaciones injustificadas de causalidad cuando se trata de simples asociaciones
- No discutir las limitaciones del diseño del estudio
- Hacer extrapolaciones y generalizaciones no justificadas
- Realizar conjeturas y comparaciones teóricas sin fundamento
- Polemizar innecesariamente y de forma trivial
- No determinar la probabilidad de cometer un error  $\beta$  (o la potencia estadística) en estudios «negativos» (estudios que no encuentran resultados estadísticamente no significativos)
- Escribir una Discusión demasiado extensa
- No hacer recomendaciones sobre futuras líneas de investigación
- Acabar la Discusión con un resumen del trabajo

lectores interesados. Estos errores se producen con más frecuencia cuando los autores copian referencias incluidas en otros trabajos sin haberlas leído.

Otros errores que se dan con frecuencia son: realizar afirmaciones sin citar la fuente que las fundamenta; apoyar conceptos ampliamente conocidos y aceptados con citas bibliográficas; omitir las citas de mayor actualidad, las españolas o las de grupos competidores, o que defienden opiniones contrarias; apoyar un concepto en una referencia que, o bien no lo desarrolla, o no lo hace en el sentido pretendido por los autores.

Es preferible evitar, en la medida de lo posible, la cita de libros de texto, de resúmenes de congresos y

reuniones científicas, la autocitación injustificada y las observaciones o comunicaciones personales no publicadas. Si se quiere hacer referencia a trabajos aceptados, pero pendientes de publicación, se incluirán en la relación bibliográfica con la cita adecuada, y añadiendo *en prensa* entre corchetes. Se recomienda incluir preferentemente artículos de los últimos 5 años y, en especial, de los últimos 2 años, y referencias fácilmente localizables.

#### RESUMEN

El resumen debe permitir a los lectores identificar la relevancia y el interés que el trabajo puede tener para ellos, y, de este modo, decidir si van a leer el artículo entero o bien optan por otro más acorde con sus necesidades. Además, el resumen es, junto con el título, el elemento en el que se basan los distintos sistemas de búsqueda bibliográfica.

Un estudio ha de comprenderse en su totalidad a partir del resumen del artículo. Este hecho es muy importante, ya que la mayoría de las bases de datos bibliográficas incluyen los resúmenes de los trabajos, pero no su texto completo.

El resumen debe redactarse cuidadosamente y debe incorporar los aspectos más importantes de cada una de las secciones del artículo: una frase con el objetivo principal del estudio; una descripción del método utilizado en dos o tres frases; resumir los resultados principales en otras dos o tres frases, y resaltar la conclusión principal del estudio. La mayoría de las publicaciones solicita que el resumen se estructure en diversos apartados para facilitar su redacción, lectura y comprensión. Los epígrafes y la extensión del resumen varían según la revista. El estilo debe ser sencillo, las frases cortas y sin desarrollar más de una idea.

Ha de repasar su contenido con minuciosidad y contrastarlo con el resto del artículo, y hay que evitar las incorrecciones y las contradicciones.

Debajo del resumen deben identificarse de tres a seis palabras clave, que ayudarán a las personas que confeccionan los índices a realizar referencias cruzadas. Es conveniente utilizar los términos del Medical Subject Headings del Index Medicus.

#### TÍTULO

El título es lo primero que se lee, y si no es atractivo o no identifica el problema que se desarrolla a continuación, muchos lectores no seguirán adelante. Además, el título puede ser la única información que aparezca en muchas bases de datos e índices, y

la difusión del artículo puede verse alterada si no refleja adecuadamente su contenido. Por estas razones, debe prestarse especial atención a la elección de las palabras que lo componen y al orden en que se escriben.

El título ha de ser explicativo, breve, claro y lo suficientemente atractivo para llamar la atención del lector cuando dé un vistazo al sumario de la revista. Su longitud no debe sobrepasar las diez palabras, y se recomienda evitar expresiones superfluas como «Un estudio sobre...», «Observaciones sobre...», «Consideración acerca de...», «Reflexión sobre...». Sin embargo, no ha de ser tan corto que sólo incluya términos muy genéricos e imprecisos. Por ejemplo, el título «Dieta y cáncer» no informa de si el autor se ha centrado en un tipo especial de cáncer o de dieta, o si ha explorado la dieta como un factor protector o de riesgo de contraer la enfermedad.

El título no debe contener nunca abreviaturas, ya que pueden tener diferentes significados según el idioma o la especialidad. Además, si una persona quiere localizar en un índice médico artículos que traten sobre la hipertensión arterial los buscará por este término, no por HTA.

## AUTORES

Paralelamente al aumento del número de revistas y de artículos médicos, se ha observado una tendencia progresiva al incremento en el número de autores. Es conocido que el número y el orden de los autores dependen, a menudo, de aspectos distintos a la contribución de cada uno de ellos en la propia investigación. La necesidad de mejorar el currículum o de adquirir más prestigio hace que, en ocasiones, algunos investigadores acuerden citarse los unos a los otros en cualquier trabajo que lleven a cabo, sin considerar el grado de implicación de cada profesional.

Ser autor es sinónimo de participación activa en todas las fases del estudio, incluida su comunicación, de tal forma que se pueda discutir y defender públicamente el contenido del artículo. Las responsabilidades que definen la autoría se hallan bien delimitadas y son ampliamente conocidas:

- Cada autor tiene que haber participado suficientemente en el trabajo representado por el artículo para asumir la responsabilidad pública de su contenido.

- La participación debe incluir tres aspectos: *a)* la concepción o el diseño del trabajo y/o el análisis y la interpretación de los datos y/o la adquisición de datos; *b)* la escritura del artículo o la revisión crítica de su contenido, y *c)* la aprobación de la versión final del texto.
- La participación única y exclusivamente en la recogida de datos no justifica la autoría.
- Cada parte del contenido del artículo que sea de valor para fundamentar las conclusiones principales, así como cada fase del trabajo, ha de poder atribuirse como mínimo a un autor.
- Las personas que han contribuido al desarrollo del trabajo, pero cuya colaboración no justifica la autoría, deben nombrarse en el apartado de agradecimientos y describirse el tipo de colaboración. Dichas personas deben dar su consentimiento para ser mencionadas.

Según las normas de Vancouver, cuando un artículo tiene múltiples autores, recomiendan citar los seis primeros e incluir al resto bajo la denominación *et al* (del latín *et alii*, que significa «y otros»), lo que da una idea de lo restrictivo del criterio de autor.

En los estudios multicéntricos en los que haya participado un grupo muy numeroso de investigadores, el grupo debe identificar a los responsables del manuscrito y dar el nombre del grupo. La National Library of Medicine indizará el nombre del grupo y el de los autores identificados como responsables de la autoría. Las revistas citarán generalmente al resto de los investigadores en el apartado de agradecimientos o en un apéndice específico.

## AGRADECIMIENTOS

En una nota al pie de la primera página o en un apéndice al texto, según los requisitos de la revista, se especificarán los agradecimientos, que por regla general, se darán por tres motivos: *a)* cualquier colaboración que no justifique la calidad de autor; *b)* la ayuda técnica, y *c)* la ayuda financiera y material; se deberá especificar su naturaleza y, en particular, mencionar las relaciones financieras que puedan dar lugar a un conflicto de intereses.

El elemento más importante de esta sección es la cortesía, y por tanto, cualquier persona cuyo nombre aparece en este apartado debe estar informada de ello y tiene que haber dado su autorización.

## BIBLIOGRAFÍA

- Aleixandre R. Bibliometría e indicadores de actividad científica. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 363-84.
- Argimon JM. Estructura: el artículo original y otros tipos de artículo. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 37-55.
- Bosch X. Escribir en inglés: normas generales e interacción con el traductor. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 109-21.
- Cortés J, Cobo E. Un paseo por las guías de publicación. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 125-44.
- García Río F. Preparativos para la redacción de un manuscrito. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 17-36.
- González JA, Jover L. El impacto de los gráficos en la comunicación científica. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 79-96.
- Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ*. 2010 Mar 23;340:c723.
- Ivers NM, Taljaard M, Dixon S, Bennett C, McRae A, Taleban J, et al. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8. *BMJ*. 2011;343:d5886.
- Jiménez Villa J. Presentación de resultados estadísticos y elaboración de tablas. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 57-77.
- Marusic A. Requisitos para el envío de un manuscrito. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 273-89.
- Mills EJ, Chan AW, Wu P, Vail A, Guyatt GH, Altman DG. Design, analysis, and presentation of crossover trials. *Trials*. 2009;10:27.
- Montgomery AA, Astin MP, Peters TJ. Reporting of factorial trials of complex interventions in community settings: a systematic review. *Trials*. 2011;12:179.
- Pulido M. El estilo científico. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M, editors. Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación. 1<sup>a</sup> ed. Barcelona: Elsevier; 2010. p. 97-107.

# Anexos

- ANEXO 1** Medidas de frecuencia **327**
- ANEXO 2** Medidas de asociación **333**
- ANEXO 3** Sensibilidad y especificidad **339**
- ANEXO 4** Análisis de la concordancia **345**
- ANEXO 5** Medidas del efecto de un tratamiento **350**
- ANEXO 6** Análisis de una revisión sistemática **353**
- ANEXO 7** Guía para la elaboración de un protocolo de estudio **358**
- ANEXO 8** Cálculo del tamaño de la muestra en situaciones especiales **361**
- ANEXO 9** Elección de la prueba estadística **364**

Página deliberadamente en blanco

La medida más básica de la frecuencia de una enfermedad es el *número* de personas que la padecen (p. ej., el pasado mes se atendieron 35 personas diagnosticadas de hipertensión arterial [HTA]). Esta medida es de gran utilidad en la planificación de los servicios sanitarios, ya que conocer cuánta gente padece una enfermedad permite distribuir mejor los recursos y adecuarlos a las necesidades de la población. Sin embargo, es de poca utilidad para la investigación de un problema de salud, ya que se desconoce el denominador o el tamaño relativo respecto de la población de la que provienen los casos. Sin un denominador es imposible comparar la frecuencia de una enfermedad en dos o más grupos de individuos.

**Ejemplo A1.1.** En la tabla A1.1 se observa que en el centro A el número de nuevos casos de HTA registrados durante un año ( $n = 50$ ) es superior al del centro B ( $n = 35$ ). Con sólo esta información se podría concluir erróneamente que la enfermedad es mucho más común entre la población atendida por el centro A. Cuando este número se relaciona con la población atendida por cada uno de los centros, resulta que la HTA es más frecuente entre la población atendida en el centro B que en el A.

### PROPORCIÓN, RAZÓN Y TASA

Una *proporción* es un cociente en el que el numerador está incluido en el denominador.

**Ejemplo A1.2.** La proporción de hipertensos esenciales controlados con dieta sin sal es de 0,75 (75%). Se trata de una proporción, ya que el numerador está formado por los hipertensos controlados con dieta, mientras que el denominador incluye a todos los hipertensos, controlados o no.

Una proporción no es más que la expresión de la probabilidad de que un suceso ocurra. En el ejemplo A1.2, corresponde a la probabilidad de que un hipertenso esté controlado con dieta sin sal. Su valor puede ir de 0 (el suceso no ocurre nunca) a 1 (todos se controlan con dieta), aunque generalmente se expresa en tanto por ciento.

Una *razón* es un cociente en el que el numerador no está incluido en el denominador. En el ejemplo A1.2, la razón entre los hipertensos controlados con dieta y los no controlados sería de 75/25 o, lo que es lo mismo, de 3/1.

Cuando en una razón el numerador representa la probabilidad de que ocurra un suceso y el denominador la probabilidad de que no ocurra, se denomina *odds*. En el ejemplo A1.2, la *odds* de control con dieta es de 3, o lo que es lo mismo, por cada hipertenso esencial que no se controla con dieta sin sal hay tres que sí lo consiguen. La escala de medida de una *odds* va desde infinito en el caso de que un suceso siempre ocurra (100/0) hasta cero cuando no ocurre nunca (0/100).

En realidad, tanto las proporciones como la *odds* expresan el mismo fenómeno, pero en formas diferentes. La relación entre ellas es la siguiente:

$$\text{proporción} = \frac{\text{odds}}{\text{odds} + 1}$$
$$\text{odds} = \frac{\text{proporción}}{1 - \text{proporción}}$$

**Ejemplo A1.3.** Sabiendo que la *odds* de los hipertensos controlados con dieta es de 3, se puede averiguar que la proporción es:  $3/(3 + 1) = 0,75$ . Igualmente, podemos convertir esta proporción en *odds*:  $0,75/(1 - 0,75) = 0,75/0,25 = 3$ .

Una *tasa* es una forma especial de proporción que tiene en cuenta el tiempo. Relaciona el cambio de una magnitud (en medicina, generalmente, es el cambio de una situación clínica) por unidad de cambio en otra magnitud (por regla general, el tiempo).

**Ejemplo A1.4.** La expresión: «la tasa anual de nuevos casos de cardiopatía isquémica entre varones de 40 a 50 años es de 8 por cada 1.000 personas en riesgo» se refiere a que, por cada aumento en la unidad de tiempo (un año), la tasa de cambio desde la situación clínica de «no enfermo» a la de «enfermo», en aquellas personas con riesgo de contraer la enfermedad (varones de 40 a 50 años), es de 8 casos por cada 1.000 personas.

Esta dimensionalidad temporal no se refiere a un período delimitado entre dos puntos temporales, sino que las tasas son medidas instantáneas del cambio del valor de una variable por cada unidad de otra variable de la cual depende la primera. Dado que para comprender exactamente el significado de una tasa instantánea se necesitan algunos conocimientos de cálculo, en investigación clínica y epidemiológica se utilizan *tasas promedio*.

**Ejemplo A1.5.** Para explicar intuitivamente el concepto de tasa promedio puede utilizarse la

**Tabla A1.1 Datos hipotéticos de la frecuencia de hipertensión arterial (HTA) en la población atendida en dos centros de salud**

Centro	Nuevos casos de HTA registrados en un año	Población	Nuevos casos de HTA registrados por 1.000 habitantes
A	50	23.500	2,13
B	35	12.200	2,87

comparación con la velocidad. Por ejemplo, la velocidad de un automóvil en un instante dado es una tasa que se expresa en kilómetros por hora. De todos modos, viajar a 80 km/h no significa necesariamente que se vayan a recorrer 80 km. En realidad, se puede ir a esta velocidad pero no conducir durante una hora. Del mismo modo, en el ejemplo anterior sobre la cardiopatía isquémica, la tasa no expresa el número de casos nuevos en un año, sino la «velocidad» con que se produce el cambio de una situación clínica a otra en una población determinada.

### PREVALENCIA

Las medidas de frecuencia de una enfermedad más comúnmente utilizadas se encuadran en dos grandes categorías: incidencia y prevalencia. La prevalencia es la proporción de individuos de una población que tiene la enfermedad:

$$\text{prevalencia} = \frac{\text{Número de personas con la enfermedad}}{\text{Población total}}$$

Normalmente se estima a partir de estudios transversales y describe la situación en un momento determinado, no lo que ocurrirá en el futuro.

**Ejemplo A1.6.** Supongamos que, del registro de los pacientes atendidos en la consulta, se extraen aquellos en los que consta el diagnóstico de HTA. Si el total de historias abiertas es de 1.000 y en 150 consta este diagnóstico, se puede afirmar que la prevalencia de HTA conocida entre los pacientes con historia abierta en la consulta es del 15% en el momento de realizar el estudio.

La prevalencia es, por definición, una proporción. Aunque coloquialmente se utiliza el término «tasa de prevalencia», en realidad no lo es.

### INCIDENCIA

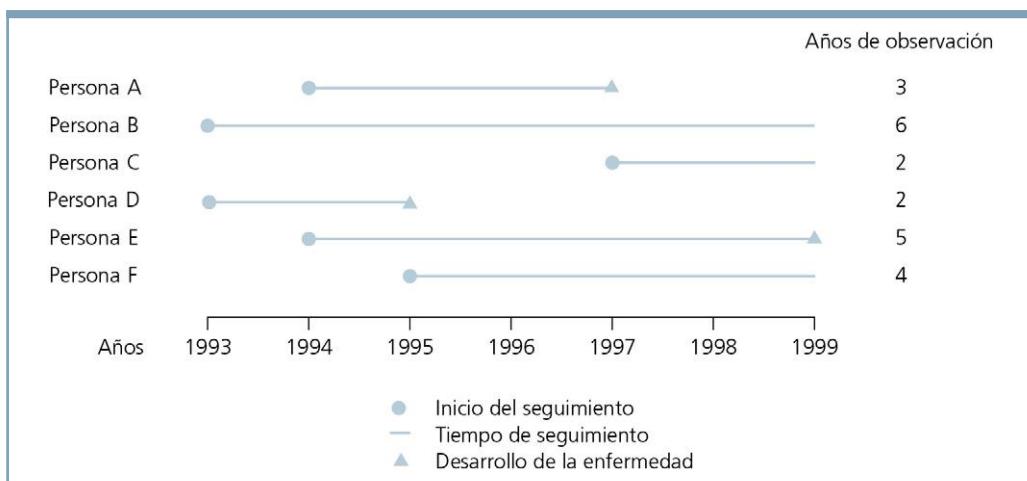
La incidencia se define como el número de casos nuevos de una enfermedad que se desarrollan en una población de riesgo durante un período determinado. Se estima a partir de estudios en los que existe un seguimiento de sujetos libres de la enfermedad, lo que permite detectar la aparición de nuevos casos. Hay dos tipos de medidas de incidencia: la incidencia acumulada y la densidad de incidencia.

La *incidencia acumulada* (IA) es la proporción de individuos sanos que a lo largo de un período determinado desarrolla la enfermedad:

$$\text{Incidencia acumulada (IA)} = \frac{\text{Número de nuevos casos en un período de tiempo}}{\text{Población de riesgo}}$$

La IA estima el riesgo, o la probabilidad, de que un individuo desarrolle la enfermedad. Es una proporción, no una tasa. Es imprescindible que se especifique el período de observación, ya que de otro modo es imposible interpretarla. Una IA del 5% puede ser muy pequeña si el período de observación ha sido de 10 años, o muy alta si la observación se ha limitado a un año. El tiempo de observación se fija arbitrariamente, pero lógicamente vendrá determinado por la naturaleza de la enfermedad que se esté estudiando.

La IA mide el riesgo de un individuo de contraer una enfermedad a condición de que no se desarrolle ninguna otra enfermedad grave que lleve a la persona a la muerte. De este modo, que un individuo de 50 años tenga un 70% de probabilidad de desarrollar cardiopatía isquémica en los próximos 25 años se interpreta como el riesgo de contraer la enfermedad siempre y cuando no muera por otra causa. Por esta razón, si se utilizan períodos de observación muy largos, la incidencia acumulada no deja de ser una medida hipotética.



**Figura A1.1** Cálculo del denominador persona-tiempo en la densidad de incidencia.

Hasta ahora se ha asumido que el tiempo de observación era el mismo para cada individuo. Sin embargo, en muchas circunstancias este tiempo es desigual, ya que las personas pueden entrar en el estudio en distintas épocas o pueden perderse a lo largo del seguimiento por causas conocidas o no.

Una posible solución sería restringir el análisis de los datos a un tiempo de observación para el cual todas las personas del estudio tuvieran información. Con ello se conseguiría igualar la duración del tiempo de observación, pero se perdería gran cantidad de información. Una estimación más precisa, y que utiliza toda la información sobre el seguimiento, es la *densidad de incidencia*, que se define como el número de nuevos casos que ocurren por unidad de tiempo-población (p. ej., años-persona) de riesgo. Se trata de una medida de la aparición de una enfermedad en un punto en el tiempo, aunque, como se ha comentado, en epidemiología se usan tasas promedio:

$$\text{Densidad de incidencia (DI)} = \frac{\text{Número de nuevos casos en un período de tiempo}}{\text{Personas - tiempo de observación}}$$

El numerador es el mismo que en la incidencia acumulada, pero el denominador no es el total de personas en situación de riesgo, sino que incorpora la dimensión «tiempo de observación».

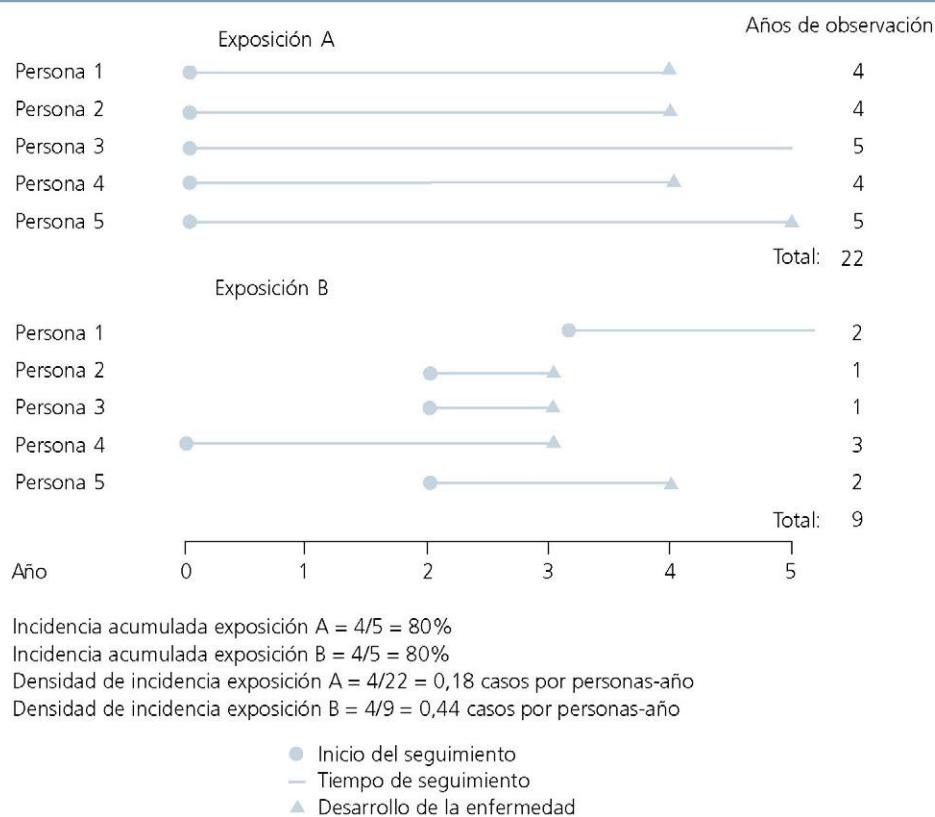
**Ejemplo A1.7.** La forma de cálculo de la densidad de incidencia se ilustra en la figura A1.1. En ella se observa a 6 personas con distintos tiempos de seguimiento, de las cuales 3 desarrollan la enfermedad de interés. La incidencia acumulada sería de 3/6 o, lo que es lo mismo, un riesgo de contraer la enfermedad del 50% en los 6 años de seguimiento.

Sin embargo, esta afirmación no tiene en cuenta el hecho de que sólo una de las 6 personas (la que corresponde a la letra B) se ha observado durante el período completo. El resto ha sido desigual: un período de 3 años en el caso de la persona A; 2 años en las personas C y D; 5 años en la persona E, y 4 años en la persona F. En total suman 22 personas-año de observación. La densidad de incidencia sería de 3 casos/22 personas-año, o lo que es lo mismo, 13,6/100 personas-año de seguimiento.

El concepto de persona-tiempo lleva implícita la idea de que el valor 10 personas-año puede proceder de la observación de 10 personas durante un año cada una, o de 5 personas durante 2 años, o en el caso más extremo de una sola persona durante 10 años. Así pues, cada unidad persona-tiempo se considera equivalente e independiente de las otras. Esta asunción es razonable, excepto en los casos, muy improbables, de que la suma de todas las unidades persona-tiempo provengan de pocos individuos (p. ej., si 30 personas-año procedieran exclusivamente de uno o dos individuos). El uso del denominador persona-tiempo es válido sólo bajo tres asunciones:

1. El riesgo de contraer la enfermedad es constante a lo largo de todo el período de observación. En ocasiones esta condición no se cumple, como es el caso de enfermedades con un largo período de latencia y en el que el seguimiento se ha llevado a cabo durante muchos años, en las que la aparición de nuevos casos tenderá a concentrarse en los últimos años de seguimiento. En esta situación es conveniente dividir el período de observación en varios subperíodos.

## Métodos de investigación clínica y epidemiológica



**Figura A1.2** Elección de la medida de incidencia.

2. La tasa de incidencia en los que se pierden durante el seguimiento es la misma que en los que continúan en observación; si esta condición no se cumple, se obtendrá un resultado sesgado.
3. El denominador es apropiado según la historia natural de la enfermedad. Si se quiere estimar la tasa de incidencia de fallos de un método anticonceptivo y se espera que en un alto porcentaje de pacientes ocurra antes de un año, no sería lógico escoger como denominador los años-persona de observación, ya que muchas personas contarían como un nuevo caso en el numerador, pero no contribuirían al denominador, por lo que la tasa de incidencia estaría artificialmente elevada. En este caso sería mejor escoger como denominador los meses-persona.

### Elección de la medida de incidencia

La elección de la medida de incidencia más adecuada viene condicionada por el objetivo

del estudio. Si el interés está en predecir el cambio en el estado de salud de un individuo en función de alguna característica, entonces se deberá estimar el riesgo de que este cambio ocurra. En este caso la densidad de incidencia será de poca ayuda, ya que no ofrece información individual.

Por otro lado, si el objetivo principal consiste en evaluar una hipótesis etiológica, la elección dependerá de la naturaleza de la enfermedad. Si ésta es crónica y con un largo período de latencia, la medida más apropiada será la densidad de incidencia. En este caso, el tiempo de seguimiento sólo representa una parte del tiempo en el que la persona tiene el riesgo de contraer la enfermedad.

**Ejemplo A1.8.** La figura A1.2 ilustra un ejemplo en el que es preferible usar la densidad de incidencia en lugar de la incidencia acumulada. Se

observa que en ambos grupos, 4 de 5 personas desarrollan la enfermedad, es decir, tienen una incidencia acumulada del 80%. Sin embargo, si nos fijamos atentamente, veremos que en el grupo con la exposición A estos 4 casos han aparecido después de 22 personas-año de observación, mientras que en el B sólo se han necesitado 9 personas-año de observación. Así, las tasas de densidad de incidencia son 0,18 personas-año ( $4/22$ ) y 0,44 personas-año ( $4/9$ ), respectivamente, lo que indica que la exposición B es mucho más agresiva (o protectora, en el supuesto de que fuera una exposición preventiva) que la A.

Si el tiempo de seguimiento incluye todo el período de riesgo, lo que sucede cuando la enfermedad y la exposición tienen períodos de latencia cortos, entonces también es de utilidad el cálculo de la incidencia acumulada para realizar inferencias etiológicas. La razón, como siempre, es mejor explicarla con un ejemplo. Supongamos que se trata de una enfermedad infecciosa, en la que el interés primordial es estimar el riesgo de un individuo de adquirirla, y si este riesgo varía según la fuente de exposición. Esto se conseguirá mejor a través de la incidencia acumulada. Hay que tener presente que la densidad de incidencia variará sustancialmente según la fase de la enfermedad, de modo que será baja al inicio, se acelerará en plena fase de propagación y volverá a disminuir al final. Una de las condiciones para usar la densidad de incidencia es que el riesgo de contraer la enfermedad sea constante a lo largo de todo el período de observación, lo que en esta situación no se cumple.

### Otras consideraciones

La existencia de un denominador preciso es una condición indispensable para poder efectuar comparaciones. Para el cálculo de la incidencia sólo deben constar en el denominador aquellas personas con riesgo de contraer la enfermedad. Por esta razón, los casos prevalentes, o bien los que por otras causas no pueden padecerla (edad, sexo, inmunización previa, etc.), deben excluirse del denominador. Si en el denominador se incluyen personas que no están en riesgo, el resultado será una incidencia disminuida. Un ejemplo claro sería calcular la tasa de incidencia de cáncer de mama incluyendo en el denominador a los hombres o mujeres muy jóvenes; en ambos casos, la posibilidad de padecer la enfermedad es muy baja, por lo que se infraestimaría la incidencia.

Sin embargo, hay que reconocer que, en las enfermedades poco frecuentes, incluir casos prevalentes en el denominador no afecta prácticamente al resultado. Así pues, la incidencia de tuberculosis en una comarca se suele calcular usando como denominador el censo de población, sin corregir por los casos prevalentes.

Por otro lado, en las estimaciones de la incidencia se ha de especificar si el numerador se refiere a casos nuevos o a episodios. En algunas enfermedades esto no es necesario, como en el caso de la diabetes, ya que no pueden ocurrir más de una vez. En otras, como el resfriado común o los ataques agudos de una enfermedad crónica (p. ej., el infarto de miocardio en la enfermedad coronaria), pueden suceder episodios en más de una ocasión, por lo que se tendrá que especificar si el numerador se refiere al número de personas que han desarrollado la enfermedad o al número de episodios.

## RELACIÓN ENTRE LA INCIDENCIA Y LA PREVALENCIA

Ambas medidas están fuertemente relacionadas. La prevalencia depende de la incidencia y de la duración de la enfermedad. Para las enfermedades cuya densidad de incidencia es relativamente estable, la prevalencia se puede expresar como el producto de la incidencia por la duración.

$$\text{Prevalencia} = \text{densidad de incidencia} \times \text{duración media de la enfermedad}$$

Por ello, es preferible el uso de casos incidentes, ya que cuando se comparan dos grupos (o dos períodos en el tiempo), una mayor prevalencia en uno de ellos puede ser debida a: *a)* una mayor duración de la enfermedad, aun cuando la incidencia pudiera ser igual o menor; *b)* una incidencia mayor con una duración similar o menor, y *c)* una incidencia y una duración mayores.

Si un factor está asociado con una enfermedad, y en el estudio se utilizan casos prevalentes, sólo cuando se den la segunda y la tercera circunstancias se podrá considerar realmente que el factor es causa de la enfermedad. Por el contrario, si se cumple la primera circunstancia, el uso de casos prevalentes llevará a la conclusión errónea de que el factor es causa de la enfermedad, cuando en realidad lo único que hace es aumentar su duración. Para evitar este problema, siempre que sea posible es conveniente utilizar casos incidentes.

## BIBLIOGRAFÍA

- Gordis L. Epidemiology. 4th ed. Philadelphia: Saunders Elsevier; 2009.
- Jager KJ, Zoccali C, Kramar R, Dekker FW. Measuring disease occurrence. *Kidney Int.* 2007;72:412-5.
- Kraemer HC. Events per person-time (incidence rate): a misleading statistic? *Stat Med.* 2009;28:1028-39.
- Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3th ed. Philadelphia: Lippincott, Williams & Wilkins; 2008.
- Szklo M, Nieto FJ. Epidemiology: beyond the basics. 2nd ed. Sudbury: Jones and Bartlett Publishers; 2007.
- Tapia Granados JA. On the terminology and dimensions of incidence. *J Clin Epidemiol.* 1997;50:891-7.

En los estudios analíticos no sólo interesa conocer si existe una asociación entre el factor de estudio y la variable de respuesta, sino también la magnitud de dicha asociación. Esto se consigue comparando la frecuencia del suceso de interés en un grupo expuesto al factor de estudio con la de un grupo no expuesto.

Para el cálculo de las medidas de asociación, los datos suelen presentarse en forma de una tabla de contingencia  $2 \times 2$  (tabla A2.1). En los estudios en los que se utiliza como medida de frecuencia la densidad de incidencia, los datos se presentan en una forma algo distinta (tabla A2.2). Estas tablas se generalizan cuando existen varios niveles de exposición.

### RIESGO RELATIVO

El riesgo relativo (RR) estima la magnitud de una asociación e indica el número de veces que es más probable que una enfermedad se desarrolle en el grupo expuesto en relación con el grupo no expuesto. En otras palabras, es la razón entre la incidencia en el grupo expuesto ( $I_e$ ) y la incidencia entre el grupo no expuesto ( $I_o$ ), por lo que también se denomina razón de incidencias o de riesgo:

$$RR = \frac{I_e}{I_o} = \frac{a/(a+b)}{c/(c+d)}$$

**Ejemplo A2.1.** Una hipotética cohorte de 368 individuos que reciben oxigenoterapia domiciliaria se ha dividido en dos grupos, según si han dejado de fumar o no. Se siguen durante un año para evaluar su mortalidad (tabla A2.3). La medida de frecuencia que se decide utilizar es la incidencia acumulada. En primer lugar, se calcula el riesgo, o probabilidad, de que un individuo que no tenga la enfermedad (en este caso, la muerte) la desarrolle durante el año de duración del estudio. El resultado es del 12,3% (19/154) en la cohorte expuesta, mientras que para los ex fumadores es del 7% (15/214). El RR de los fumadores respecto a los ex fumadores es de  $12,3/7,0 = 1,8$ , lo que significa que, en un año, un fumador tiene 1,8 veces más probabilidades de morir que un ex fumador.

Cuando la medida de frecuencia es la densidad de incidencia, el RR se calcula de forma similar:

$$RR = \frac{I_e}{I_o} = \frac{a/PT_e}{c/PT_o}$$

**Ejemplo A2.2.** Supongamos que la tasa de incidencia de cardiopatía isquémica es de 13,0 por

1.000 personas-año en hombres fumadores de 50 a 65 años, mientras que en los no fumadores de este mismo grupo de edad sólo es de 6,2 por 1.000 personas-año de observación. El RR es el cociente entre ambas tasas de incidencia:  $13,0/6,2 = 2,1$ . Este resultado se interpreta como que la tasa de aparición de nuevos casos de cardiopatía isquémica en fumadores es 2,1 veces la que se observa en no fumadores.

Un RR de 1,0 indica que no existe relación entre el factor de estudio y la enfermedad. Si es mayor que 1,0 indica que existe una asociación positiva entre el factor de estudio y la enfermedad. Cuando es menor que 1,0 indica una asociación negativa.

El RR obtenido en un estudio es una estimación puntual y, por lo tanto, debe calcularse su intervalo de confianza (IC). Si el IC del 95% no incluye el valor  $RR = 1$ , existe una asociación estadísticamente significativa entre el factor de estudio y el desenlace. Para calcularlo, debe determinarse, en primer lugar, el error estándar del logaritmo natural (neperiano) del RR:

$$EE(\ln RR) = \sqrt{\frac{b}{(a+b) \times a} + \frac{d}{(c+d) \times c}}$$

Donde  $\ln RR$  es el logaritmo natural del RR. El IC de  $\ln RR$  se calcula del siguiente modo:

$$\ln RR \pm Z_{\alpha/2} \times EE(\ln RR)$$

Dado que la fórmula anterior permite obtener los límites del IC del logaritmo del RR, deberán determinarse sus antilogaritmos para conocer el IC del RR.

**Ejemplo A2.3.** En un estudio hipotético se ha seguido una cohorte de 500 sujetos expuestos y otra de 500 sujetos no expuestos a un factor de riesgo, y se han observado los resultados de la tabla A2.4. La estimación puntual del RR es 4 y los límites de su IC del 95% son 2,1 y 7,4. Dado que el IC excluye el valor 1, existe una asociación estadísticamente significativa entre la exposición y la enfermedad.

### ODDS RATIO

En los estudios de cohortes, el RR se estima directamente, ya que se conoce la incidencia de la enfermedad en los individuos expuestos y no expuestos a un factor. Por el contrario, en los estudios de casos y controles no se puede calcular la incidencia, porque la población de estudio se selecciona

## Métodos de investigación clínica y epidemiológica

**Tabla A2.1** Presentación de una tabla  $2 \times 2$  para el cálculo de las medidas de asociación

	Enfermedad	No enfermedad	Total
Exposición	a	b	a+b
No exposición	c	d	c+d
Total	a+c	b+d	a+b+c+d

Las casillas a, b, c y d representan diferentes combinaciones entre la enfermedad y la exposición:

a: número de personas expuestas que tienen la enfermedad.

b: número de personas expuestas que no tienen la enfermedad.

c: número de personas no expuestas que tienen la enfermedad.

d: número de personas no expuestas que no tienen la enfermedad.

Los marginales de las tablas resultan de sumar las correspondientes casillas:

a+b: número total de expuestos.

c+d: número total de no expuestos.

a+c: número total de personas con la enfermedad.

b+d: número total de personas sin la enfermedad.

La suma de las cuatro casillas, a+b+c+d, representa el número total de participantes en el estudio.

**Tabla A2.2** Presentación de una tabla para el cálculo de las medidas de asociación cuando el denominador son unidades de persona-tiempo de observación

	Enfermedad	Persona-tiempo
Exposición	a	PTe
No exposición	c	PTo
Total	a+c	PTe + PTo

a: número de personas enfermas en el grupo expuesto.

c: número de personas enfermas en el grupo no expuesto.

PTe: unidades de persona-tiempo de observación en el grupo expuesto.

PTo: unidades de persona-tiempo de observación en el grupo no expuesto.

En consecuencia, los marginales serán:

a+c: total de personas enfermas.

PTe + PTo: total de unidades de persona-tiempo de observación.

**Tabla A2.3** Resultados de un estudio hipotético de una cohorte de 368 individuos que reciben oxigenoterapia domiciliaria (ejemplo A2.1)

	Número de defunciones	Número de pacientes que sobreviven	Total
Fumadores	19	135	154
Exfumadores	15	199	214
Total	34	334	368

Incidencia acumulada en el grupo expuesto:  $le = 19/154 = 12,3\%$ .

Incidencia acumulada en el grupo no expuesto:  $lo = 15/214 = 7,0\%$ .

Riesgo relativo:  $RR = 12,3/7,0 = 1,8$ .

Diferencia de incidencias:  $DI = 12,3 - 7,0 = 5,3\%$ .

**Tabla A2.4 Ejemplo de cálculo del intervalo de confianza (IC) del riesgo relativo (RR)**

	<b>Enfermos</b>	<b>No enfermos</b>	<b>Total</b>
Expuestos	48	452	500
No expuestos	12	488	500
Total	60	940	1.000
Estimación puntual del RR: $RR = \frac{48/500}{12/500} = 4$			
Error estándar del logaritmo del RR:			
$EE(\ln RR) = \sqrt{\frac{452}{500 \times 48} + \frac{488}{500 \times 12}} = 0,31$			
IC del 95% del $\ln RR$ :			
$\ln 4 \pm 1,96 \times 0,31$			
$1,39 \pm 0,62$			
$0,77 \times 2,01$			
Límites del IC del 95% del RR:			
Límite inferior: $\exp(0,77) = 2,15$			
Límite superior: $\exp(2,01) = 7,44$			

a partir de individuos que ya han desarrollado la enfermedad. La razón  $a/(a+b)$ , por ejemplo, no estima el riesgo de contraer la enfermedad en aquellos expuestos al factor de riesgo, sino que refleja, en parte, decisiones tales como elegir dos controles por caso.

La medida de asociación que se utiliza en los estudios de casos y controles es la razón de *odds* o *odds ratio* (OR). Por *odds* se entiende la razón entre la probabilidad de que un suceso ocurra y la probabilidad de que no ocurra. Si la probabilidad de que una persona con la enfermedad esté expuesta es del 0,75, la *odds* de exposición se calculará dividiendo este valor por la probabilidad de no estar expuesto ( $0,75/[1 - 0,75] = 3$ ). La OR no es más que la razón entre la *odds* de exposición observada en el grupo de casos ( $a/c$ ) y la *odds* de exposición en el grupo control ( $b/d$ ).

$$OR = \frac{(a/c)}{(b/d)} = \frac{a \times d}{b \times c}$$

Para explicar intuitivamente de dónde proviene la fórmula de la OR partiremos del esquema de un estudio de cohortes. Supongamos que toda una población se clasifica según la presencia o la ausencia de un factor de riesgo, y que se sigue durante un tiempo para observar en cuántos individuos aparece la enfermedad y clasificarlos de acuerdo con la combinación exposición-enfermedad que presenten. El RR sería:

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

En la mayoría de las enfermedades, la proporción de individuos clasificados como enfermos será pequeña, es decir,  $a$  será pequeño en comparación con  $b$ , por lo que el total de personas expuestas  $a+b$  será prácticamente igual a  $b$ , y lo mismo sucede con  $c$  en comparación con  $d$ , por lo que el total de personas no expuestas  $c+d$  será igual a  $d$ . Por lo tanto, la anterior ecuación sería aproximadamente:

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{a/b}{c/d} = \frac{a \times d}{c \times b}$$

De este modo, un estudio de casos y controles puede conceptualizarse como un estudio en el que el grupo de casos lo forman todos los individuos que tienen la enfermedad ( $a+c$ ), y los controles se escogen entre aquellos que no la tienen ( $b+d$ ). Lógicamente no se estudian todos los individuos, sino una muestra de ellos. Si los casos y los controles se han elegido independientemente de la historia de exposición previa, la OR puede considerarse una buena estimación del RR. El ejemplo de la tabla A2.5 ilustra este concepto.

La OR aventaja al RR en que su resultado es independiente de si se expresa en términos de riesgo de contraer la enfermedad o de no contraerla.

**Ejemplo A2.4.** En el estudio hipotético de la tabla A2.6, el RR de enfermar en la comunidad A respecto al RR en la comunidad B es  $(2/100)/(1/100) = 2$ . La OR es  $(2/98)/(1/99) = 2$ . Así pues, ambas medidas dan el mismo resultado. Consideremos la

## Métodos de investigación clínica y epidemiológica

**Tabla A2.5 Ejemplo que ilustra por qué la odds ratio (OR) es una buena estimación del riesgo relativo (RR)**

a. Estudio de cohortes hipotético diseñado para evaluar el riesgo de úlcera gástrica asociada a la ingesta de alcohol

	Enfermos	No enfermos	Total
Alcohol ≥ 80 g/día	175	2.825	3.000
Alcohol < 80 g/día	207	6.793	7.000
RR = 1,97			

b. Estudio de casos y controles hipotético realizado en la misma población. Los casos son todos los individuos que tienen la enfermedad y, como controles, una muestra (10%) de los no enfermos

	Casos	Controles	Total
Alcohol ≥ 80 g/día	175	282	457
Alcohol < 80 g/día	207	679	886
OR = 2,03			

**Tabla A2.6 Comparación del riesgo de enfermar en dos comunidades (ejemplo A2.4)**

Comunidad	Enfermos	No enfermos	Total
A	2	98	100
B	1	99	100
Total	3	197	200

posibilidad de analizar los datos desde la perspectiva de no enfermar. El RR sería  $(98/100)/(99/100) = 1$ , es decir, no habría ninguna asociación entre vivir en una comunidad y el hecho de no enfermar. La OR sería  $(98/2)/(99/1) = 1/2$ . Con el RR se obtiene un resultado distinto, según se exprese en relación con el hecho de enfermar o con el de no enfermar. Sin embargo, con la OR, el riesgo de enfermar en la comunidad A es el doble que en la comunidad B, mientras que la probabilidad de no enfermar en esta última comunidad es la mitad. Es decir, proporciona el mismo resultado independientemente del fenómeno que se exprese, lo que siempre es deseable en una medida de asociación.

La estimación puntual de la OR debe acompañarse de su correspondiente intervalo de confianza (IC). Un método sencillo y aproximado para su cálculo consiste en calcular el error estándar (EE) del logaritmo natural (o neperiano) de la OR:

$$EE(\ln OR) = \sqrt{(1/a + 1/b + 1/c + 1/d)}$$

Donde  $\ln OR$  es el logaritmo natural de la OR. El IC de  $\ln OR$  se calcula del siguiente modo:

$$\ln OR \pm Z_{\alpha/2} \times EE(\ln OR)$$

La fórmula permite obtener los límites del IC del logaritmo de la OR. Para conocer los límites del IC de la OR se deberán determinar sus respectivos antilogaritmos.

**Ejemplo A2.5.** Supongamos un estudio con 200 casos y 200 controles en los que se evalúa una exposición, en el que se obtienen los resultados de la tabla A2.7. La estimación puntual de la OR es 4,4 y los límites del IC del 95% son 2,6 y 7,4. La asociación entre la exposición y la enfermedad es estadísticamente significativa, ya que el IC del 95% de la OR excluye el valor 1.

La OR se utiliza cada vez con más frecuencia como medida de la magnitud de un efecto o asociación, tanto en los estudios de casos y controles, de cohortes y ensayos clínicos, como en revisiones sistemáticas y metaanálisis.

**Tabla A2.7 Ejemplo de cálculo del intervalo de confianza (IC) de la odds ratio (OR)**

	Casos	Controles	Total
Expuestos	70	22	92
No expuestos	130	178	308
Total	200	200	400

Estimación puntual de la OR:  $OR = (70 \times 178) / (22 \times 130) = 4,36$

Error estándar del logaritmo de la OR:  
 $ES(\ln OR) = \sqrt{1/70 + 1/22 + 1/130 + 1/178} = \sqrt{0,073} = 0,27$

IC del 95% del  $\ln OR$ :  
 $\ln 4,36 \pm 1,96 \times 0,27$   
 $1,47 \pm 0,53$   
 0,94 a 2,00

IC del 95% de la OR:  
 Límite inferior:  $\exp(0,94) = 2,56$   
 Límite superior:  $\exp(2,00) = 7,39$

Habitualmente se interpreta como equivalente del RR, aunque en realidad la OR sólo es una buena aproximación del RR bajo determinadas condiciones. Cuando el riesgo basal (frecuencia del resultado de interés en la población de estudio) es bajo (menor del 10%), ambas medidas son muy similares. Cuanto mayor es el riesgo basal, la OR más exagera la magnitud del efecto en relación con el RR, es decir, más lo sobreestima cuando es mayor que 1 y más lo infraestima cuando es menor de 1. De hecho, la diferencia entre la OR y el RR depende tanto del riesgo basal como de la propia magnitud del efecto, de forma que grandes discrepancias sólo se observan en efectos de gran tamaño y riesgos basales elevados. Por ello, en la gran mayoría de las ocasiones, la interpretación de la OR como RR no conduce a cambios en la interpretación cualitativa de los resultados, aunque debe tenerse en cuenta que

puede existir cierta sobreestimación de la magnitud del efecto.

En los estudios de cohortes suele utilizarse la regresión logística para controlar múltiples factores de confusión, obteniéndose una OR ajustada. Zhang y Yu (1998) proponen una fórmula para corregir este valor y obtener una mejor estimación del RR:

$$RR = \frac{OR}{(1-P_o) + (P_o \times OR)}$$

Donde  $P_o$  es la frecuencia del resultado de interés en el grupo no expuesto (riesgo basal). Esta misma corrección puede aplicarse a los límites del IC.

Las medidas de asociación no miden el impacto potencial sobre la salud de la población de la exposición al factor de estudio o de su eliminación. Las medidas de impacto se presentan en el capítulo 32.

## BIBLIOGRAFÍA

- Altman DG. Confidence interval for odds ratio. *Physiother Res Int.* 2000;5:134-5.
- Bland JM, Altman DG. The odds ratio. *BMJ.* 2000;320:1468.
- Davies HTO, Crombie IK, Tavakoli M. When can odds ratio mislead? *BMJ.* 1998; 316:989-91.
- Gordis L. *Epidemiology.* 4th ed. Philadelphia: Saunders Elsevier; 2009.
- Grimes DA, Schulz KF. Making sense of odds and odds ratios. *Obstet Gynecol.* 2008;111:423-6.
- Katz KA. The (relative) risks of using odds ratios. *Arch Dermatol.* 2006;142:761-4.
- Martínez González MA, De Irala Estévez J, Guillén Grima F. ¿Qué es una odds ratio? *Med Clin (Barc).* 1999;112: 416-22.
- Nurminen M. To use or not to use the odds ratio in epidemiologic analyses? *Eur J Epidemiol.* 1995;11:365-71.

## Métodos de investigación clínica y epidemiológica

- Pearce N. What does the Odds Ratio estimate in a case-control study? *Int J Epidemiol.* 1993;22:1189-92.
- Prasad K, Jaeschke R, Wyer P, Keitz S, Guyatt G. Evidence-Based Medicine Teaching Tips Working Group. Tips for teachers of evidence-based
- medicine: understanding odds ratios and their relationship to risk ratios. *J Gen Intern Med.* 2008;23:635-40.
- Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology.* 3th ed. Philadelphia: Lippincott: Williams & Wilkins; 2008.
- Szklo M, Nieto FJ. *Epidemiology: beyond the basics.* 2nd ed. Sudbury: Jones and Bartlett Publishers; 2007.
- Zhang J, Yu KF. What's the relative risk? A method of correcting the Odds Ratio in cohort studies of common outcomes. *JAMA.* 1998;280:1690-1.

## ANEXO 3. Sensibilidad y especificidad

El equipo investigador debe valorar la validez de las medidas y seleccionar la más adecuada para su estudio. La sensibilidad, la especificidad y el valor predictivo son los criterios de validez que cuantifican la capacidad de una prueba para clasificar correcta o erróneamente a una persona, según la presencia o ausencia de una exposición o una enfermedad. Las pruebas diagnósticas son imperfectas y se cometen errores al clasificar a una persona por su resultado.

### SENSIBILIDAD Y ESPECIFICIDAD

La validez de una medida se calcula partiendo de la información contenida en una tabla de  $2 \times 2$  (tabla A3.1). La presencia o ausencia del resultado (en general una enfermedad) se determina a partir de un criterio de referencia, que idealmente debe ser siempre positivo en los individuos con la enfermedad y negativo en aquellos que no la presentan. Por otro lado, se encuentra el resultado de la prueba, o medida en general, que se quiere evaluar. Para simplificar la exposición, la mayor parte de las veces se considerará que se trata de una medida dicotómica, por lo que su resultado se clasificará en positivo o negativo.

De los resultados obtenidos en cada una de las casillas de la tabla A3.1 se derivan, entre otros,

dos índices: la sensibilidad y la especificidad. La *sensibilidad* responde a la pregunta: si un individuo tiene una enfermedad o factor de riesgo, ¿qué probabilidad existe de que el resultado de la medida que se le aplica sea positivo? En otras palabras, la probabilidad de que una medida clasifique correctamente a un individuo enfermo. La *especificidad* responde a la pregunta: si un individuo no tiene la enfermedad o el factor de riesgo, ¿qué probabilidad existe de que el resultado obtenido sea negativo? Representa la probabilidad de que una medida clasifique correctamente a una persona no enferma. Ambos conceptos son características intrínsecas de la propia medida. Es decir, que si ésta se aplica en una población de individuos semejante, y los observadores tienen la misma experiencia, la sensibilidad y la especificidad de una medida no variarán cuando se emplee en distintos estudios.

**Ejemplo A3.1.** Con el objetivo de evaluar un método simple y económico de tinción directa para el diagnóstico de infección urinaria, en un estudio se analizaron 1.064 muestras de orina (García Curiel, 1988). Se utilizó como colorante el azul de metileno, que permite teñir las bacterias y el material nuclear de leucocitos. Esta prueba se comparó con el método tradicional de siembra en placa de agar (criterio de referencia). Las

**Tabla A3.1** Presentación de resultados de un estudio de valoración de una prueba diagnóstica

Resultados de la prueba	Clasificación de los individuos según el criterio de referencia		
	Enfermos	No enfermos	Total
Positivo	a	b	$n_1$
Negativo	c	d	$n_2$
Total	$m_1$	$m_2$	N

Verdaderos positivos (a): número de individuos con la enfermedad, en los que el resultado de la prueba diagnóstica es positivo.

Falsos positivos (b): número de individuos sin la enfermedad, en los que el resultado de la prueba diagnóstica es positivo.

Falsos negativos (c): número de individuos con la enfermedad, en los que el resultado de la prueba diagnóstica es negativo.

Verdaderos negativos (d): número de individuos sin la enfermedad, en los que el resultado de la prueba diagnóstica es negativo.

De esta información se derivan los siguientes índices:

Sensibilidad ( $a/m_1$ ): probabilidad de obtener un resultado positivo en los individuos que tienen la enfermedad.

Especificidad ( $d/m_2$ ): probabilidad de obtener un resultado negativo en los individuos que no tienen la enfermedad.

Valor predictivo positivo ( $a/n_1$ ): probabilidad de que un individuo que presenta un resultado de la prueba positivo tenga la enfermedad.

Valor predictivo negativo ( $a/n_2$ ): probabilidad de que un individuo que presenta un resultado de la prueba negativo no tenga la enfermedad.

orinas se clasificaron como positivas si existía un crecimiento de  $10 - 10^5$  UFC/ml (unidades formadoras de colonias). Si el crecimiento era inferior a  $10^5$  UFC/ml se consideró que no existía bacteriuria significativa.

Los resultados principales se exponen en la tabla A3.2. Se encontraron 346 bacteriurias, lo que representa el 32,5% del total de las muestras estudiadas. La sensibilidad de la tinción para detectar una infección urinaria fue del 82,4% (284/346), y la especificidad fue del 92,3% (706/718). Así pues, en el grupo de orinas con bacteriuria significativa se identificó con la nueva medida un 82,4%. De igual modo, el resultado fue negativo en un 98,3% de las muestras consideradas negativas. La tinción fue falsamente negativa en un 17,6% de los casos (100 – sensibilidad) y sólo fue falsamente positiva en un 1,7% de las ocasiones (100 – especificidad).

Variando el criterio de normalidad se modifican los valores de estos parámetros. Si en el ejemplo A3.1 se considerara que un paciente presenta bacteriuria cuando los valores fueran superiores a  $10^4$  UFC/ml, se aumentaría la sensibilidad de la medida a costa de reducir su especificidad. Del mismo modo, si se fuese más exigente en el criterio de normalidad, aumentaría la especificidad y disminuiría la sensibilidad.

### VALORES PREDICTIVOS

En la práctica clínica, cuando un médico solicita una prueba diagnóstica desconoce si el paciente tiene la enfermedad. Los médicos deben efectuar inferencias sobre la presencia o ausencia de la enfermedad a partir de los resultados de la prueba. Existen dos modos de cuantificar esta inferencia: los valores predictivos y los cocientes de probabilidad.

El *valor predictivo positivo* (VPP) es la probabilidad de que un individuo con un resultado positivo tenga la enfermedad, y el *valor predictivo negativo* (VPN) es la probabilidad de que si el resultado es

negativo el paciente no tenga la enfermedad. Los valores predictivos dependen no sólo de la sensibilidad y la especificidad, sino también de la prevalencia de la enfermedad.

**Ejemplo A3.2** En el ejemplo A3.1 (ver tabla A3.2), el VPP es del 96% (285/297) y el VPN es del 92% (706/767). El trabajo se llevó a cabo en pacientes ingresados en un hospital o que acudían a las consultas externas, y la prevalencia fue del 32,5%. Si se aplicara la misma prueba a la población que consulta en un centro de salud, la prevalencia de infección urinaria sería, muy probablemente, menor y, por consiguiente, el VPP disminuiría. Supongamos que la prevalencia es del 5% y se aplica la prueba a 1.000 personas, la sensibilidad y la especificidad serían las mismas que las calculadas en la tabla A3.2. Los resultados de este estudio hipotético se muestran en la tabla A3.3. La predictividad de la prueba positiva ha disminuido de un 96% a un 71,9%.

La predictividad de una medida no se puede evaluar sin considerar la prevalencia de la enfermedad; si es alta, un resultado positivo tiende a confirmar su presencia, mientras que si es negativo, no ayudará a excluirla. Contrariamente, cuando la prevalencia es baja, un resultado negativo permitirá descartar la enfermedad con un elevado margen de confianza, pero si es positivo, no permitirá afirmar su existencia.

La prevalencia es el factor más determinante de los valores predictivos. La sensibilidad y la especificidad, al ser características intrínsecas de una medida, no sufrirán grandes variaciones según el lugar donde se apliquen, siempre y cuando se realicen en condiciones similares. Sin embargo, esta asunción no siempre se cumple. El espectro de pacientes también varía según el lugar donde se aplica la prueba. Por ejemplo, una misma prueba, cuando se emplea en un programa de detección precoz se aplica a sujetos asintomáticos, mientras que cuando se usa con fines diagnósticos en un hospital de alta tecnología, muchos de los pacientes que la reciben tienen una enfermedad avanzada.

**Tabla A3.2 Resultados de un estudio que evalúa un nuevo método diagnóstico de infección urinaria**

	Cultivo positivo	Cultivo negativo	Total
Tinción positiva	285	12	297
Tinción negativa	61	706	767
Total	346	718	1.064

**Tabla A3.3** Resultados de un estudio que evalúa un nuevo método diagnóstico de infección urinaria

	Infección urinaria	No infección urinaria	Total
Tinción positiva	41	16	57
Tinción negativa	9	934	943
Total	50	950	1.000

Valor predictivo positivo:  $41/57 = 71,9\%$ .  
 Valor predictivo negativo:  $934/943 = 99,0\%$ .  
 \*Los valores de sensibilidad (82,4%) y especificidad (98,3%) son los mismos que los calculados en la tabla A3.2.

Dado que muchos de los pacientes que son vistos en atención primaria con un resultado positivo en una prueba diagnóstica (es decir, la suma de verdaderos y falsos positivos) son derivados a otro nivel de atención, es de esperar que la especificidad se reduzca.

**Ejemplo A3.3.** En una muestra de 2.000 pacientes visitados en atención primaria con un diagnóstico de sospecha de apendicitis aguda, los que tenían una elevada probabilidad de padecerla eran derivados al hospital de referencia para confirmación y tratamiento. Una comparación entre los resultados observados en ambos niveles de atención mostró que la prevalencia de la enfermedad fue del 14% en las consultas de atención primaria y del 63% entre los pacientes derivados a los servicios de urgencia de los hospitales. Este aumento de prevalencia se acompañó también de una distinta prevalencia de los síntomas y signos diagnósticos. Así, el dolor en el cuadrante inferior derecho se observó en el 21% de los pacientes visitados en atención primaria, mientras que en los derivados al hospital esta prevalencia fue del 82%. La derivación de pacientes con resultados falsos positivos supuso que la especificidad del signo dolor en el cuadrante inferior derecho disminuyera desde el 89% hasta el 16%. Como consecuencia, un signo diagnóstico útil en atención primaria (cociente de probabilidad de una prueba positiva de 8 y cociente de probabilidad de una prueba negativa de 0,2) carece de utilidad en los hospitales de referencia (cociente de probabilidad de una prueba positiva y negativa de 1) (Sackett DL, 2002).

La especificidad no siempre se reduce cuando los pacientes son derivados al nivel secundario o terciario de atención, por lo que no existe un «factor» que sirva para ajustar los resultados en función del nivel de atención. La única forma de evitar este problema es repitiendo el mismo estudio en distintas poblaciones y lugares.

## RAZONES DE PROBABILIDAD

La razón o cociente de probabilidad compara la probabilidad de obtener un determinado resultado en un individuo que presente la enfermedad con la de obtenerlo en un sujeto en el que se ha descartado su presencia.

La razón de probabilidad de una prueba positiva (RPP) se calcula dividiendo la proporción de casos que tiene un resultado positivo de la prueba (sensibilidad), entre la proporción de personas que no tienen la enfermedad, pero en las que la prueba también ha dado un resultado positivo ( $1 -$ especificidad):

$$RPP = \frac{\text{sensibilidad}}{1 - \text{especificidad}}$$

**Ejemplo A3.4.** En el ejemplo A3.1 (ver tabla A3.2), la RPP sería  $82,4\%/(100\% - 98,3\%) = 49$ , lo cual permitiría interpretar que en el grupo de pacientes diagnosticados de infección urinaria la probabilidad de encontrar un resultado positivo a la tinción con azul de metileno es 49 veces mayor que en los individuos en los que se ha descartado la enfermedad.

Análogamente, la razón de probabilidad de una prueba negativa (RPN) se calcula dividiendo los casos que tienen un resultado negativo ( $1 -$ sensibilidad) entre la proporción de sujetos que no tienen la enfermedad, en los que el resultado de la prueba es negativo (especificidad):

$$RPN = \frac{1 - \text{sensibilidad}}{\text{especificidad}}$$

**Ejemplo A3.5.** Siguiendo con el mismo ejemplo anterior, la RPN sería  $(100\% - 82,4\%)/98,3 = 0,18$ , lo que indica que un resultado negativo se encontró 5,5 veces ( $1/0,18 = 5,5$ ) más frecuentemente en los individuos sin infección urinaria que entre aquellos que si la padecieron (ver tabla A3.2).

La razón de probabilidad relaciona la sensibilidad y la especificidad en un solo índice, por lo que no varía con la prevalencia. Pueden obtenerse razones de probabilidad según varios valores de una nueva medida y no es necesario expresar la información de forma dicotómica, como resultado normal o anormal, o positivo y negativo.

A partir de la razón de probabilidad se pueden calcular los valores predictivos (o probabilidad a posteriori) de una prueba. El primer paso consiste en expresar la prevalencia en *odds* de enfermedad. Si, por ejemplo, la prevalencia es del 25%, la *odds* será de 1:3 (25%/75%). A continuación se multiplica la *odds* de la enfermedad por la RPP, obteniendo así la *odds posprueba*.

$$\text{Odds posprueba} = \text{odds preprueba} \times \text{RPP}$$

Esta *odds posprueba* se puede convertir en *probabilidad posprueba*, o lo que es lo mismo, en los valores predictivos, de la siguiente forma:

$$\text{probabilidad posprueba} = \frac{\text{odds posprueba}}{(\text{odds posprueba} + 1)}$$

**Ejemplo A3.6.** En el ejemplo anterior, la prevalencia era del 32,5%, o lo que es lo mismo, una *odds* preprueba de 1:2, y la RPP era de 49. Por lo tanto, la *odds posprueba* = 1:2 × 49 = 49:2 y la probabilidad posprueba = 24,5/(24,5 + 1) = 0,96. Es decir, el VPP es del 96%, el mismo que el obtenido anteriormente.

Una prueba útil desde el punto de vista clínico es aquella que, una vez aplicada, genera cambios desde la estimación diagnóstica de la probabilidad preprueba hasta la estimación de la probabilidad posprueba. Desde el punto de vista clínico, las pruebas proporcionan una ganancia de información cuando la prevalencia de la enfermedad es intermedia, superior al 10%. Las pruebas con una RPP superior a 10 o una RPN inferior a 0,1 tienen una gran utilidad clínica; si los valores de las razones de probabilidad están comprendidos entre 5 y 10 o 0,1 y 0,2, tienen una utilidad moderada.

### CURVAS ROC

Cuando los valores de la prueba diagnóstica siguen una escala cuantitativa, la sensibilidad y la especificidad varían según el punto de corte elegido para clasificar a la población como enferma o no enferma. En esta situación, una forma más global de conocer la exactitud de una prueba en el conjunto de puntos de corte es mediante el uso de curvas

ROC (*receiver operating characteristics* o curvas de características operativas para el receptor).

La curva ROC es un gráfico en el que en el eje de ordenadas se sitúa la sensibilidad (proporción de verdaderos positivos) y en el eje de abscisas el complementario de la especificidad (1 – especificidad o proporción de falsos positivos) (fig. A3.1). Cada punto de la curva representa el valor de la RPP correspondiente a un punto de corte determinado.

El área bajo la curva se define como la probabilidad de clasificar correctamente a un par de individuos, uno sano y otro enfermo, seleccionados al azar, al aplicarles la prueba, y es independiente de la prevalencia de la enfermedad. Por ejemplo, un área bajo la curva de 0,75 significa que un individuo seleccionado aleatoriamente del grupo de enfermos tendrá el 75% de las veces un valor de la prueba mayor que un individuo sano elegido al azar. No significa que un resultado positivo se produzca en los individuos enfermos con una probabilidad de 0,75 ni que esté asociado con la enfermedad el 75% de las veces.

Para una prueba con una sensibilidad y una especificidad del 100%, la curva ROC estaría representada por los lados izquierdo y superior de la figura (área bajo la curva igual a 1). Cuando la prueba no tiene ningún poder de discriminación, es decir, se observan los mismos resultados en los individuos enfermos que en los sanos, la curva ROC está representada por la diagonal principal del gráfico (área bajo la curva igual a 0,5). Como norma general, si el área bajo la curva es mayor de 0,9, la prueba es muy exacta, mientras que valores comprendidos entre 0,7 y 0,9 indican una exactitud moderada. Los valores comprendidos entre 0,5 y 0,7 se corresponden con una exactitud baja.

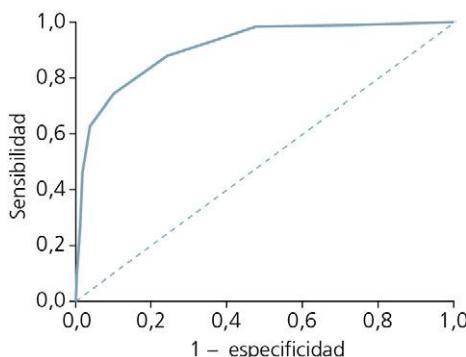


Figura A3.1 Ejemplo de curva ROC.

En definitiva, cuanto más próxima es una curva ROC a la esquina superior izquierda, más alta es la exactitud global de la prueba.

La curva ROC facilita la elección del punto de corte. La mayoría de las curvas ROC presentan un segmento de gran pendiente, en el cual la sensibilidad aumenta mucho, mientras la proporción de falsos positivos prácticamente no varía. No tiene mucho sentido elegir el punto de corte en esta zona, porque al desplazarse hacia arriba por la curva aumentará la sensibilidad sin reducirse sustancialmente la especificidad. Igualmente, tampoco es aconsejable elegir el punto de corte en la zona plana de la curva, ya que la sensibilidad se mantiene prácticamente inalterada mientras que la proporción de falsos positivos aumenta. Por lo general, si el coste que supone cometer un falso positivo es similar al de cometer un falso negativo, el mejor punto de corte es el más próximo al ángulo superior izquierdo del gráfico. Si el coste de un falso positivo y uno negativo difieren, se deberá tener en cuenta este distinto coste para el cálculo del punto óptimo de corte.

Los intervalos de confianza del área bajo la curva permiten efectuar comparaciones estadísticas entre distintas pruebas diagnósticas, siempre que se apliquen en la misma población y para responder a una misma duda diagnóstica. Cuando los resultados de las pruebas se expresan de forma dicotómica, se pueden comparar mediante la *odds ratio* diagnóstica, que se calcula del siguiente modo:

$$OR \text{ diagnóstica} = \frac{\text{sensibilidad} \times \text{especificidad}}{(1-\text{sensibilidad}) \times (1-\text{especificidad})}$$

Se puede demostrar matemáticamente que la *odds ratio* diagnóstica es independiente de los cambios que ocurren en la prevalencia y del espectro de enfermedad que presentan los pacientes. Los investigadores pueden usar técnicas multivariantes, como la regresión logística, para identificar el punto de corte que optimiza el resultado de la *odds ratio*. Las pruebas útiles suelen tener *odds ratio* superiores a 20 (es decir, una RPP de 7 y una RPN de 0,3).

## BIBLIOGRAFÍA DE LOS EJEMPLOS

- García Curiel A. Diagnóstico precoz de las infecciones del tracto urinario: examen microscópico y cuantitativo de orina total teñida con azul de metileno. *Enf Infec Microbiol Clin.* 1988;6: 303-7.
- Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ.* 2002;324: 539-41.

## BIBLIOGRAFÍA

- Akobeng AK. Understanding diagnostic tests 2: likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta Paediatr.* 2007;96:487-91.
- Altman DG. ROC curves and confidence intervals: getting them right. *Heart.* 2000;83:236.
- Barratt A, Irwig L, Glasziou P, Cumming RG, Raffle A, Hicks N, et al. Users' guides to the medical literature: XVII. How to use guidelines and recommendations about screening. Evidence-Based Medicine Working Group. *JAMA.* 1999;281:2029-34.
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou P, Irwig L, et al. Standards for Reporting of Diagnostic Accuracy. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med.* 2003;138(1):W1-12.
- Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol.* 2003;56:1129-35.
- Grimes DA, Schulz KF. Uses and abuses of screening tests. *Lancet.* 2002;359:881-4.
- Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. Designing studies to ensure that estimates of test accuracy are transferable. *BMJ.* 2002;324:669-71.
- Langlotz CP. Fundamental measures of diagnostic examination performance: usefulness for clinical decision making and research. *Radiology.* 2003;228:3-9.
- Lu Y, Heller DN, Zhao S. Receiver operating characteristic (ROC) analysis for diagnostic examinations with uninterpretable cases. *Stat Med.* 2002;21:1849-65.
- Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med.* 2002;137: 598-602.

## Métodos de investigación clínica y epidemiológica

Richardson WS, Wilson MC,  
Keitz SA, Wyer PC. EBM  
Teaching Scripts Working  
Group. Tips for teachers of  
evidence-based medicine:  
making sense of diagnostic  
test results using likelihood  
ratios. *J Gen Intern Med.*  
2008;23:87-92.

Scales Jr CD, Dahm P, Sultan  
S, Campbell-Scherer D,  
Devereaux PJ. How to use an  
article about a diagnostic test. *J  
Urol.* 2008;180:469-76.  
Tze-Wey L. Understanding  
sensitivity and specificity with  
the right side of the brain.  
*BMJ.* 2003;327:716-9.

van Stralen KJ, Stel VS, Reitsma  
JB, Dekker FW, Zoccali C,  
Jager KJ. Diagnostic methods  
I: sensitivity, specificity, and  
other measures of accuracy.  
*Kidney Int.* 2009;75:1257-63.

## ANEXO 4. Análisis de la concordancia

El análisis de la concordancia entre diferentes medidas de una variable permite evaluar tanto la repetibilidad como la variabilidad intraobservador e interobservador, así como comparar diversos métodos de medida de una misma variable. Las técnicas de análisis dependen del tipo de variable.

### VARIABLES CUALITATIVAS: ÍNDICE KAPPA

Una forma sencilla de expresar la concordancia cuando la variable es cualitativa es la proporción o el porcentaje observado de mediciones concordantes. Sin embargo, dado que una parte de esta concordancia se debe al azar, es necesario utilizar otros métodos que tengan en cuenta este hecho.

El índice Kappa resume la concordancia entre dos medidas de una variable (p. ej., las obtenidas por dos observadores), cuando está en una escala cualitativa, eliminando la fracción de la concordancia debida al azar, es decir, la que se obtendría si las dos medidas no estuvieran relacionadas. En la tabla A4.1 se muestra la presentación general de los datos para su cálculo en el caso de una medida dicotómica. La fórmula es:

$$Kappa = \frac{P_o - P_e}{1 - P_e}$$

Donde  $P_o$  es la proporción total de concordancia observada y  $P_e$  la proporción de concordancia esperada por azar.

**Ejemplo A4.1.** En un estudio hipotético, dos médicos leen 100 radiografías y las clasifican como normales o anormales. Las evaluaciones se realizan de forma independiente, es decir, ninguno de los médicos conoce el resultado de su colega. El médico 1 clasifica 30 radiografías como anormales, y el médico 2, 35, y ambos coinciden en 20 casos. Los resultados se muestran en la tabla A4.2. La proporción total observada de concordancia entre ambos médicos ( $P_o$ ) es:

$$P_o = \frac{a + d}{N} = \frac{20 + 55}{100} = 0,75 \text{ o bien } 75\%$$

Esta proporción observada no tiene en cuenta el papel del azar y se pueden obtener conclusiones erróneas. Supongamos que observar una radiografía anormal fuera muy poco frecuente. El número de coincidencias negativas sería probablemente muy elevado, por lo que contribuiría de una manera desproporcionada al porcentaje total de concordancias. Un observador, incluso sin mirar las radiografías, simplemente diciendo que todas son normales, podría ser concordante con el resultado del otro observador en un porcentaje elevado de casos. El porcentaje total de concordancia o aciertos no es, pues, un buen índice. Dado que se calcula a partir de la suma de coincidencias sobre la positividad o negatividad de un hallazgo, un índice del 0,75 se puede lograr igualmente a partir de una suma del 0,40 más el 0,35, o de sumar 0,70 más 0,05, por ejemplo, un resultado este último mucho más fácil de conseguir.

**Tabla A4.1 Presentación de una tabla 2 × 2 para el cálculo de los índices de concordancia**

Número de hallazgos del observador 1	Número de hallazgos del observador 2		
	Positivos	Negativos	Total
Positivos	a	b	$n_1$
Negativos	c	d	$n_2$
Total	$m_1$	$m_2$	N

a: número de coincidencias de hallazgos positivos entre ambos observadores.

b: número de veces que el observador 1 califica un hallazgo de positivo y el observador 2 de negativo.

c: número de veces que el observador 1 califica un hallazgo de negativo y el observador 2 de positivo.

d: número de coincidencias de hallazgos negativos entre ambos observadores.

$n_1$ : número total de hallazgos positivos del observador 1.

$n_2$ : número total de hallazgos negativos del observador 1.

$m_1$ : número total de hallazgos positivos del observador 2.

$m_2$ : número total de hallazgos negativos del observador 2.

N: número total de observaciones.

**Tabla A4.2 Concordancia entre dos observadores que leen 100 radiografías (ejemplo A4.1)**

Hallazgos observador 2	Hallazgos observador 1		<b>Total</b>
	Anormales	Normales	
Radiografías			
Anormales	20	15	35
Normales	10	55	65
Total	30	70	100

El grado de concordancia esperable por azar puede determinarse a partir de los valores marginales de la tabla de contingencia. Así, para calcular en cuántas ocasiones los observadores concordarían por azar en valorar las radiografías como anormales, se multiplicaría el número de pruebas que cada médico hubiese clasificado como anormales y se dividiría por el número total de observaciones:

$$\frac{(a + b) \times (a + c)}{N} = \frac{30 \times 35}{100} = 10,5$$

De forma similar, para calcular en cuántas ocasiones los observadores concordarían por azar en valorar las radiografías como normales:

$$\frac{(c + d) \times (b + d)}{N} = \frac{70 \times 65}{100} = 45,5$$

Sumando estos resultados y dividiendo entre el total, se obtiene la proporción esperada de concordancia por azar:

$$P_e = \frac{10,5 + 45,5}{100} = 0,56$$

Como la proporción total observada  $P_o$  es 0,75, el índice Kappa será:

$$Kappa = \frac{P_o - P_e}{1 - P_e} = \frac{0,75 - 0,56}{1 - 0,56} = 0,43$$

El valor máximo de un índice Kappa es 1, que se produce cuando existe un acuerdo total entre ambos observadores. El valor mínimo no es cero, sino que puede ser negativo. Si el porcentaje observado es menor que el esperado por azar, el índice Kappa tendrá un valor negativo. Este hecho ha sido criticado por algunos autores, aunque en realidad tiene poca trascendencia, ya que cuando el porcentaje observado es menor que el esperado, la concordancia es tan baja que no interesa saber su valor exacto.

La concordancia se considera muy débil cuando los valores del índice Kappa son inferiores a 0,20, débil si está entre 0,21 y 0,40, moderada entre 0,41 y 0,60, buena entre 0,61 y 0,80, y muy buena si es superior a 0,80.

La fórmula permite obtener una estimación puntual del índice Kappa. Un método sencillo de calcular su intervalo de confianza es el siguiente:

$$IC\ 95\%: Kappa \pm 1,96 \times \sqrt{\frac{P_o(1 - P_o)}{N(1 - P_o)}}$$

El índice Kappa se propuso originalmente como un índice de concordancia entre dos observadores para clasificar variables dicotómicas. Posteriormente, se propuso su uso para variables politómicas. Sin embargo, cuando se usa en esta situación, tiene el inconveniente de que mide la frecuencia del acuerdo exacto, más que la del acuerdo aproximado. Es decir, no tiene en cuenta el grado de desacuerdo, ya que todos los desacuerdos son tratados de igual manera. Una alternativa es calcular el índice Kappa ponderado, en el que se asignan diferentes ponderaciones a los desacuerdos, según su magnitud. Cuantas más categorías tenga la variable, más disminuye la probabilidad de acuerdo. De hecho, cuando se acerca a una escala cuantitativa, el concepto de probabilidad de acuerdo se convierte en un absurdo. En realidad, cuando se trata de variables ordinales, muchos autores recomiendan utilizar el coeficiente de correlación intraclass en lugar del índice Kappa ponderado. De todas formas, con variables politómicas puede ser más informativo el uso de varios índices Kappa para diferentes combinaciones entre las categorías de la variable que un único valor global promedio.

El índice Kappa no es el mejor índice para resumir la validez de una medida. Siempre que se posea un buen criterio de referencia, es preferible usar la sensibilidad y la especificidad para valorar la validez de una medida. Si la concordancia es elevada, existe la posibilidad, aunque no la garantía, de que las mediciones reflejen la dimensión que se pretende medir (validez). Pero si la concordancia es baja, la utilidad de las mediciones será muy limitada, ya que no tiene sentido preguntarse si están asociadas

con la variable que se desea medir si uno no se puede fiar ni de las propias mediciones.

## VARIABLES CUANTITATIVAS

En el caso de variables cuantitativas, existe cierta confusión entre los conceptos de relación, o asociación, y de concordancia, de forma que es frecuente el uso de técnicas estadísticas inadecuadas.

**Ejemplo A4.2.** La diferencia entre relación y concordancia se ilustra con los datos presentados en la tabla A4.3, que corresponden a un estudio hipotético para determinar la concordancia entre una técnica estándar de medición de la presión arterial sistólica (PAS) y tres nuevos métodos A, B y C. Puede observarse que todos ellos tienen valores del coeficiente de correlación de Pearson con el método estándar de 1, de modo que indican una relación lineal perfecta. Sin embargo, también puede apreciarse que los valores no coinciden en todos los casos. Mientras que el método A presenta una correlación y una concordancia perfectas con la técnica estándar, el método B proporciona valo-

res consistentemente superiores en 30 mmHg, lo que se refleja en la diferencia entre las medias de las mediciones. En cambio, aunque el método C tiene una correlación perfecta y la media de sus mediciones coincide con la del método estándar, existe una discordancia evidente en los valores que proporcionan para cada sujeto. Así pues, si bien una concordancia perfecta implica una correlación perfecta, lo contrario no se cumple necesariamente, incluso aunque los valores medios sean similares.

Del ejemplo se deduce que el coeficiente de correlación de Pearson no es una medida adecuada del grado de concordancia entre dos mediciones, sino que mide el grado de relación lineal entre ellas, de forma que, por ejemplo, si una es sistemáticamente el doble de la otra, la correlación será perfecta. Tampoco la regresión lineal es una técnica adecuada, ya que se obtiene una ecuación que permite predecir el valor de una variable en función de otra, pero no indica el grado de acuerdo entre ellas.

Como puede deducirse fácilmente de los datos de la tabla A4.3, la comparación de medias tampoco es adecuada para realizar este análisis. Si se

**Tabla A4.3 Estudio hipotético que compara la medición de la presión arterial sistólica (PAS) con la técnica estándar y tres nuevos métodos**

Sujeto	Técnica estándar		Método A		Método B		Método C					
	PAS	PAS	Dif.	PAS	Dif.	PAS	Dif.					
1	100	100	0	130	+ 30	120	+ 20					
2	110	110	0	140	+ 30	125	+ 15					
3	120	120	0	150	+ 30	130	+ 10					
4	130	130	0	160	+ 30	135	+ 5					
5	140	140	0	170	+ 30	140	0					
6	150	150	0	180	+ 30	145	- 5					
7	160	160	0	190	+ 30	150	- 10					
8	170	170	0	200	+ 30	155	- 15					
9	180	180	0	210	+ 30	160	- 20					
Media	140	140	0	170	+ 30	140	0					
			r = 1 R = 1			r = 1 R = 0,625						
	Dif.: diferencia respecto al valor obtenido por la técnica estándar. r: coeficiente de correlación de Pearson con el método estándar. R: coeficiente de correlación intraclass con el método estándar.											

utilizara como criterio la igualdad de las medias de las mediciones, podría concluirse que tanto el método A como el C tienen una buena concordancia con el método estándar, lo que, evidentemente, es erróneo en lo que se refiere al método C. Además, al aplicar una prueba estadística para comparar las medias, se hacen sinónimos la concordancia y la ausencia de significación estadística. Por este criterio, cuanto mayor sea el error de medida, es decir, la variabilidad de los valores, menor será la probabilidad de encontrar diferencias estadísticamente significativas, y, por lo tanto, mejor será la concordancia, lo que resulta un error evidente.

Una forma sencilla de cuantificar la concordancia es determinar el tanto por ciento de sujetos en los que el nuevo método proporciona un valor cercano al obtenido por el de referencia (p. ej.,  $\pm 5$  mmHg en la toma de la presión arterial). Sin embargo, si bien este porcentaje puede proporcionar una información útil, ignora el hecho de que una determinada fracción de la concordancia puede deberse simplemente a la variabilidad aleatoria, incluso aunque ambos métodos de medida sean totalmente independientes. Por ello, se hace necesario un índice estadístico que cuantifique la concordancia corrigiendo por este fenómeno.

### Coeficiente de correlación intraclass

El coeficiente de correlación intraclass (R) es un índice que sintetiza el grado de concordancia entre dos variables cuantitativas. Su definición matemática deriva del modelo del análisis de la varianza. La variabilidad total de las mediciones ( $\sigma^2_x$ ) se puede descomponer, por un lado, en la debida a los verdaderos valores ( $\sigma^2_v$ ), es decir, la de las diferencias entre sujetos, y, por otro, a la variabilidad residual de los errores de medida ( $\sigma^2_e$ ), es decir, la de las diferencias entre las medidas para cada sujeto. Cuanto menor sea esta variabilidad residual en relación con la total, mayor será la concordancia entre las mediciones:

$$R = \frac{\sigma^2_v}{\sigma^2_v + \sigma^2_e} = \frac{\sigma^2_v}{\sigma^2_x}$$

Los valores de R varían entre 0 y 1. El valor 0 corresponde a la situación en que el componente residual supone toda la variabilidad. A medida que constituye una porción menor de la misma, el valor de R aumenta, hasta alcanzar el valor máximo de 1 cuando  $\sigma^2_e = 0$ . La concordancia suele considerarse muy buena si R es mayor de 0,90, buena si está entre 0,71 y 0,90, moderada entre 0,51 y 0,70, mediocre entre 0,31 y 0,50, y mala o nula si R es inferior a 0,31.

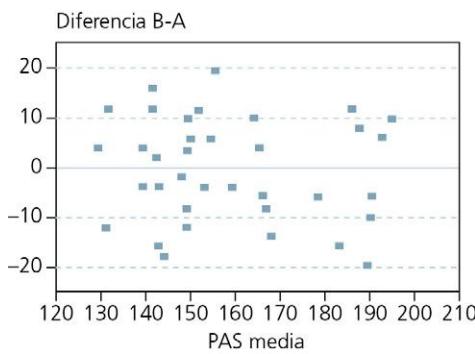
En la tabla A4.3 puede apreciarse cómo R refleja el grado de concordancia mejor que el coeficiente de correlación de Pearson.

### Análisis de las diferencias individuales

Bland y Altman (1986) proponen un método sencillo a partir del análisis de las diferencias individuales, que proporciona información complementaria al coeficiente de correlación intraclass.

El primer paso es la representación gráfica de las diferencias entre ambos métodos en relación con su media (fig. A4.1). La diferencia de los resultados entre ambos métodos se representa en el eje de ordenadas, y su promedio, en el de abscisas. Lo ideal sería representar las diferencias observadas en relación con el verdadero valor de la variable de interés, lo que puede hacerse cuando uno de los métodos puede considerarse el estándar de referencia (se estaría en una situación de evaluación de la validez de una prueba o medida). Sin embargo, cuando se evalúa la concordancia, se asume que ninguno de los métodos puede considerarse de referencia y se utiliza la media de ambos métodos como la mejor estimación disponible.

Si la concordancia fuera perfecta, los puntos se situarían exactamente sobre la línea correspondiente a la diferencia igual a 0. Este gráfico también permite observar si la magnitud de las diferencias entre los métodos es similar para todos los valores de la variable (representados por los promedios de ambas mediciones), o si las diferencias entre los métodos son mayores a medida que aumenta el valor de la variable.



**Figura A4.1** Diferencias en los valores de presión arterial sistólica (PAS) entre dos métodos de medida A y B en relación con su promedio.

Si las diferencias no varían según los valores de la variable, pueden analizarse las diferencias individuales: su media aritmética estima el sesgo relativo (diferencia sistemática entre ambos métodos) y su desviación estándar, el error aleatorio. Los límites del intervalo de dos desviaciones estándar alrededor de la media de las diferencias pueden considerarse como los límites de concordancia entre ambas mediciones, ya que este intervalo incluye el 95% de las diferencias observadas.

Este análisis de las diferencias también puede aplicarse a la evaluación de la repetibilidad. Para ello, basta con representar las diferencias entre ambas observaciones en el eje de ordenadas, y su

promedio en el de abscisas, y realizar el análisis como se ha descrito.

El método de Bland y Altman se centra en la valoración de la magnitud de la diferencia, ya que permite determinar los límites de concordancia observados. Estos valores deben compararse con los límites de concordancia establecidos antes del inicio del estudio, que dependerán de la utilidad que se quiera dar a la medida. Este método considera la concordancia más como un concepto clínico que como uno estadístico. De esta forma, la determinación de la magnitud de las diferencias y su visualización gráfica facilitan su interpretación clínica y hacen que este método sea muy atractivo para los profesionales sanitarios.

## BIBLIOGRAFÍA

- Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol.* 2003;22:85-93.
- Bland JM, Altman DG. Measurement error. *BMJ.* 1996;312:1654.
- Bland JM, Altman DG. Measurement error and correlation coefficients. *BMJ.* 1996;313:41-2.
- Bland JM, Altman DG. Measurement error proportional to the mean. *BMJ.* 1996;313:106.
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8: 135-60.
- Bland JM, Altman DG. Statistical Methods for Assessing Agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-10.
- Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol.* 1993;46:423-9.
- Graham P, Jackson R. The analysis of ordinal agreement data: beyond weighted kappa. *J Clin Epidemiol.* 1993;46:1055-62.
- Guggenmoos-Holzmann I. The meaning of kappa: probabilistic concepts of reliability and validity revisited. *J Clin Epidemiol.* 1996;49:775-82.
- Haber M, Gao J, Barnhart HX. Assessing observer agreement in studies involving replicated binary observations. *J Biopharm Stat.* 2007;17:757-66.
- Hamilton C, Stamey J. Using Bland-Altman to assess agreement between two medical devices don't forget the confidence intervals! *J Clin Monit Comput.* 2007;21:331-3.
- Kundel HL, Polansky M. Measurement of observer agreement. *Radiology.* 2003;228:303-8.
- Latour J, Abraira V, Cabello JB, López Sánchez J. Las mediciones clínicas en cardiología: validez y errores de medición. *Rev Esp Cardiol.* 1997;50:117-28.
- Luiz RR, Costa AJ, Kale PL, Werneck GL. Assessment of agreement of a quantitative variable: new graphical approach. *J Clin Epidemiol.* 2003;56:963-7.
- MacLure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol.* 1987;126:161-9.
- Mandrekar JN. Measures of interrater agreement. *J Thorac Oncol.* 2011;6:6-7.
- Van Stralen KJ, Jager KJ, Zoccali C, Dekker FW. Agreement between methods. *Kidney Int.* 2008;74:1116-20.

## ANEXO 5. Medidas del efecto de un tratamiento

El diseño más adecuado para evaluar la eficacia de un tratamiento o una intervención es el ensayo clínico aleatorio (ECA). Típicamente, en este tipo de estudios, una muestra de pacientes es distribuida aleatoriamente en dos grupos, uno de los cuales recibe el tratamiento en estudio y el otro, la intervención de referencia (otro tratamiento o un placebo), y posteriormente se comparan las respuestas observadas en cada uno de los grupos. Para poder valorar la relevancia clínica de los resultados es necesario determinar la magnitud de la diferencia observada entre los grupos, lo que puede realizarse con medidas tanto absolutas como relativas.

**Ejemplo A5.1.** Supongamos un ECA que compara un betabloqueante con un placebo, con el fin de evaluar su eficacia en la reducción de la mortalidad en la insuficiencia cardíaca congestiva después de 2 años de tratamiento (tabla A5.1). En el grupo que ha recibido el placebo, al finalizar el estudio un 12,5% de los sujetos ha fallecido, mientras que sólo lo ha hecho un 7,5% de los que han recibido el tratamiento.

Una forma sencilla de expresar el resultado es mediante la diferencia de incidencias acumuladas

en ambos grupos, que se conoce como reducción absoluta de riesgo (RAR):

$$RAR: Io - It = 0,125 - 0,075 = 0,05 \text{ (5\%)}$$

Significa que el tratamiento ha reducido en un 5% la mortalidad a los 2 años en sujetos con insuficiencia cardíaca, en relación con el placebo. Dado que la RAR corresponde a una diferencia de proporciones, puede calcularse su intervalo de confianza (IC) del 95% utilizando la fórmula:

$$IC \ 95\% \quad (Io - It) \pm Z\alpha \times EED$$

Donde EED es el error estándar de la diferencia:

$$EED = \sqrt{\frac{Io(1-Io)}{n_o} + \frac{It(1-It)}{n_t}}$$

$$EED = \sqrt{\frac{0,125 \times 0,875}{1200} + \frac{0,075 \times 0,925}{1200}}$$

$$= 0,012$$

Luego, el IC del 95% del RAR es:

$$IC \ 95\% \ RAR: 0,05 \pm 1,96 \times 0,012$$

$$IC \ 95\% \ RAR: 5\% \pm 2,4\%$$

Es decir, que los límites del IC son del 2,6 y el 7,4%, lo que indica que existe un 95% de confianza

**Tabla A5.1** **Medidas del efecto de un tratamiento. Resultados de un ensayo clínico aleatorio hipotético que compara un tratamiento activo con placebo para reducir la mortalidad en dos grupos de 1.200 sujetos con insuficiencia cardíaca congestiva**

		Defunción		
		Sí	No	Total
Intervención	Tratamiento	90 a	1.110 b	1.200 a+b
	Placebo	c 150	d 1.050	c+d 1.200
	Total	a+c 240	b+d 2.160	N 2.400

Riesgo (incidencia) en el grupo tratado:  $It = a/(a+b) = 90/1.200 = 0,075 \text{ (7,5\%)}$

Riesgo (incidencia) en el grupo placebo:  $Io = c/(c+d) = 150/1.200 = 0,125 \text{ (12,5\%)}$

Reducción absoluta de riesgo (\*):  $RAR = Io - It = 0,125 - 0,075 = 0,05 \text{ (5\%)}$

Riesgo relativo del tratamiento respecto a placebo:  $RR = It/Io = 0,075/0,125 = 0,6$

Riesgo relativo del placebo respecto al tratamiento:  $RR = Io/It = 0,125/0,075 = 1,67$

Reducción relativa de riesgo (\*\*):  $RRR = (Io - It)/Io = (0,125 - 0,075)/0,125 = 0,4$

$RRR = 1 - RR = 1 - 0,6 = 0,4$

Número de personas que es necesario tratar para prevenir un caso adicional de insuficiencia cardíaca:

$$NNT = 1/RAR = 1/(0,125 - 0,075) = 20$$

(\*) Cuando se comparan incidencias de un suceso deseable, corresponde al aumento absoluto de beneficio:  $AAB = It - Io$

(\*\*) Cuando se comparan incidencias de un suceso deseable, corresponde al aumento relativo de beneficio:

$$ARB = (It - Io)/Io$$

de que la verdadera diferencia entre los grupos está situada entre dichos valores.

Otra forma de expresar los resultados sería como riesgo relativo (RR) (anexo 2) del grupo tratado en relación con el de comparación, que corresponde al cociente entre las incidencias observadas en cada grupo. En el ejemplo, su valor sería  $0,075/0,125 = 0,6$ , lo que significa que la mortalidad en 2 años en el grupo tratado es 0,6 veces la observada en el grupo que ha recibido placebo. Dado que este valor es inferior a 1, indicaría que ha habido una disminución del riesgo (en este caso, de la mortalidad) en el grupo tratado en relación con el grupo de placebo. Su IC 95% puede calcularse por el procedimiento descrito en el anexo 2, de modo que se obtendría un intervalo cuyos límites son 0,47 y 0,77. Dado que el valor RR = 1 queda excluido del intervalo, la reducción del riesgo observada es estadísticamente significativa.

También podría calcularse el RR del grupo placebo en relación con el tratado, que sería  $0,125/0,075 = 1,67$ , lo que significa que el riesgo de mortalidad a los 2 años es 1,67 veces superior en los sujetos que han recibido placebo en relación con los que han recibido el tratamiento. En este caso, los límites del IC 95% son 1,30 y 2,14.

Los resultados también pueden presentarse como la reducción relativa de riesgo (RRR) que se consigue con el tratamiento, es decir, la reducción de la incidencia en un grupo, expresado en forma de tanto por ciento respecto al otro grupo:

$$RRR = \frac{Io - It}{Io} = \frac{0,125 - 0,075}{0,125} = 0,4$$

Es decir, en el grupo tratado, la mortalidad a los 2 años es un 40% menor que la observada en el grupo placebo.

Obsérvese que el valor 0,4 es el complementario del RR del grupo tratado en relación con el placebo, que era de 0,6 ( $1 - 0,6 = 0,4$ ), por lo que la RRR puede expresarse también como:

$$RRR = \frac{Io - It}{Io} = 1 - \frac{It}{Io} = 1 - RR$$

Los límites del IC 95% de la RRR son el 23 y el 53%, lo que indica que se tiene un 95% de confianza de que la verdadera RRR está situada entre dichos valores.

La aplicación práctica de los ensayos clínicos puede ser difícil para el profesional sanitario, que debe valorar en un determinado individuo los riesgos y los beneficios de la instauración de un tratamiento farmacológico a largo plazo. Otra

medida que se puede obtener de estos estudios, fácil de calcular a partir de la RAR, comprensible para el clínico y extrapolable a la realidad de su labor diaria, es el número de casos que es necesario tratar (NNT) para evitar un suceso. Desde el punto de vista matemático, esta medida equivale al recíproco del valor de la reducción absoluta de riesgo y, por tanto, presenta las mismas ventajas en la expresión de resultados:

$$NNT = \frac{1}{RAR} = \frac{1}{Io - It}$$

Presenta la ventaja adicional de que indica el esfuerzo terapéutico que debe realizarse para poder prevenir un episodio.

**Ejemplo A5.2.** En el ejemplo anterior, el número de personas que es necesario tratar (NNT) para prevenir una defunción es:

$$NNT = \frac{1}{0,125 - 0,075} = \frac{1}{0,05} = 20$$

Esta cifra indica que es necesario tratar a 20 personas de características similares a las del estudio con el betabloqueante durante 2 años para prevenir una defunción más que si se hubiera administrado un placebo.

Una forma sencilla de calcular su IC 95% es calcular el valor del NNT para cada uno de los límites del IC 95% del RAR, que en el ejemplo eran del 2,6 y el 7,4%. Por lo tanto, los límites del IC 95% del NNT serían sus valores recíprocos, es decir, 13,5 y 38,5 (o bien, 14 y 39).

Un cálculo similar al del NNT puede aplicarse para estimar el número de personas que es necesario cribar para detectar un caso de enfermedad o prevenir una muerte, o incluso el número de personas que es necesario tratar para que aparezca un efecto secundario. El cálculo de estas medidas permite hacer un balance entre los riesgos y los beneficios esperados del tratamiento.

La expresión de los resultados en términos relativos (como el RR o la RRR), como a menudo se hace al difundir las conclusiones de los ensayos clínicos, puede hacer que se sobrevalore el efecto beneficioso real del fármaco. En el ejemplo, si el resultado se presenta como una RRR del 40% de la mortalidad, parece reflejar un efecto del tratamiento mucho mayor que si se presenta como RAR (reducción de la mortalidad en un 5%). Sin embargo, ambas medidas representan los mismos resultados, aunque desde perspectivas diferentes. Por tanto, la valoración adecuada de la magnitud del efecto del fármaco requiere ambos tipos de medidas.

Por otro lado, las medidas absolutas que permiten evaluar la utilidad práctica de un tratamiento, como la RAR o el NNT, varían en función del riesgo basal de la población a la que se aplican, tal como se discute más ampliamente en el capítulo 32.

**Ejemplo A5.3.** Como se ha visto en el ejemplo anterior, el tratamiento tiene una RRR, en relación con el placebo, del 40%. Si se administra en una población de alto riesgo (p. ej., con una Io = 50%), la reducción del 40% de este riesgo supone una RAR más elevada, del 16% (que corresponde a un NNT

de alrededor de 6), de modo que el riesgo del grupo tratado se reduce a una It = 34%. Si la población tiene un riesgo muy bajo (p. ej., con una Io = 5%), esta misma reducción relativa supone una disminución de la incidencia en el grupo tratado (It) al 3,4%, es decir, un RAR del 1,6% (que corresponde a un NNT de 62). Por tanto, para valorar adecuadamente la magnitud del efecto de un tratamiento, hay que conocer, además de la RRR, la incidencia real del proceso que se intenta reducir en la población a la que se va a aplicar, lo que permitirá determinar el valor de las medidas absolutas en dicha población.

## BIBLIOGRAFÍA

- Akobeng AK. Understanding measures of treatment effect in clinical trials. *Arch Dis Child.* 2005;90:54-6.
- Austin PC. Absolute risk reductions, relative risks, relative risk reductions, and numbers needed to treat can be obtained from a logistic regression model. *J Clin Epidemiol.* 2010;63:2-6.
- Citrome L. Relative vs. absolute measures of benefit and risk: what's the difference? *Acta Psychiatr Scand.* 2010;121:94-102.
- Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ.* 1995;310:452-4.
- Hutton JL. Number needed to treat and number needed to harm are not the best way to report and assess the results of randomised clinical trials. *Br J Haematol.* 2009;146:27-30.
- Rembold CM. Number needed to screen: development of a statistic for disease screening. *BMJ.* 1998;317:307-12.
- Salmi LR, Suissa S, Chêne G, Salamon R. Clinically informative measures of the effect of drugs or other interventions. *Br J Clin Pharmacol.* 2008;65:935-41.
- Tripepi G, Jager KJ, Dekker FW, Wanner C, Zoccali C. Measures of effect: relative risks, odds ratios, risk difference, and 'number needed to treat'. *Kidney Int.* 2007;72:789-91.
- Wen L, Badgett R, Cornell J. Number needed to treat: a descriptor for weighing therapeutic options. *Am J Health Syst Pharm.* 2005;62:2031-6.

A continuación se desarrollan algunos aspectos de la estrategia de análisis de una revisión sistemática (RS), cuyo esquema básico se ha presentado en el capítulo 10. La finalidad del análisis es obtener, si es adecuado hacerlo, un estimador combinado del efecto de un factor de estudio sobre una variable de respuesta a partir de los resultados obtenidos en diferentes estudios. El procedimiento estadístico que se utiliza para ello se denomina *metaanálisis*, que en términos generales, consta de dos fases: en la primera se calcula, para cada estudio, un estadístico que cuantifique el efecto de la intervención, y estos estadísticos se combinan en una segunda fase en forma de un promedio ponderado.

A continuación se comentan las principales fases del análisis de una RS.

### ANÁLISIS DE LA HETEROGENEIDAD

La principal condición necesaria para combinar los resultados de diferentes estudios es que las diferencias entre ellos sean debidas exclusivamente al azar, es decir, que se trate de estudios homogéneos. Por ello, la primera fase del análisis es evaluar si se cumple esta condición.

La calidad de los resultados del análisis depende, por una parte, de la calidad de los datos de los estudios individuales y, por otra, de las diferencias entre ellos (características de la población estudiada, definiciones de las condiciones clínicas, características del factor de estudio, variables de respuesta analizadas, etc.). Por tanto, el primer paso consiste en presentar todos estos datos en una tabla, denominada *tabla de evidencias*, para poder evaluar si es razonable combinar los resultados de los estudios.

La heterogeneidad de los resultados de los estudios individuales no se refiere exclusivamente a la situación en que no van en la misma dirección (p. ej., cuando algunos estudios sugieren que una determinada exposición es protectora frente al desarrollo de una enfermedad, mientras que otros estudios sugieren que es un factor de riesgo de la misma), sino también cuando las estimaciones de cada estudio son de magnitud muy diferente, aunque no necesariamente en direcciones opuestas (p. ej., cuando algunos estudios sugieren que una exposición aumenta ligeramente el riesgo de una enfermedad, mientras que otros sugieren que el aumento de riesgo es muy importante).

Dado que los estudios que se analizan son una muestra de todos los posibles estudios que pueden realizarse sobre un tema, sus resultados pueden

diferir simplemente por azar. Por un lado, aun cuando el efecto fuera el mismo en todos los estudios, los resultados observados variarían alrededor de este «efecto común fijo» (variabilidad intraestudio). Además, dado que los estudios pueden presentar ciertas diferencias en la población, en las características de la enfermedad o en las de la intervención o factor de estudio, es de esperar que puedan existir también ciertas diferencias en la estimación del efecto (variabilidad interestudios).

La presencia de esta heterogeneidad puede evaluarse mediante pruebas estadísticas. Si se pone de manifiesto la existencia de heterogeneidad, es obligado analizar sus posibles causas y, en ocasiones, abstenerse de combinar los resultados. Sin embargo, si el resultado de la prueba no es estadísticamente significativo, no puede descartarse completamente la existencia de heterogeneidad, especialmente si el número de estudios es reducido y con muestras de pequeño tamaño, ya que en esta situación existe gran variabilidad y la potencia de la prueba para detectar diferencias es pequeña. Por ello, se sugiere que, en lugar de usar el nivel de significación estadística habitual del 0,05, se use el 0,10, con el fin de paliar parcialmente este problema.

El análisis estadístico debe complementarse con una representación gráfica conjunta de los resultados de los estudios individuales y sus correspondientes intervalos de confianza (IC), lo que permitirá la inspección visual de la magnitud de las diferencias. Si se detecta la presencia de una clara heterogeneidad, la mejor estrategia es explorar sus posibles causas (p. ej., mediante análisis de subgrupos o una metaregresión) y no combinar sus resultados.

### ESTIMACIÓN COMBINADA DEL EFECTO

Si se considera razonable combinar los resultados de los estudios, se realiza un metaanálisis para obtener una estimación conjunta del efecto. Su utilidad principal es que, al incluir en el análisis un mayor número de sujetos que en los estudios individuales, aumenta la potencia y la precisión estadísticas, de modo que se obtienen IC más estrechos. También puede ser útil para responder preguntas no abordadas por los estudios individuales o para explorar aparentes controversias entre los resultados de los estudios y generar nuevas hipótesis. Obviamente, su uso no garantiza la validez de los resultados de la RS, ya que ésta depende

fundamentalmente de la calidad de los estudios individuales. A veces puede ser conveniente no realizar un único análisis, sino varios combinando conjuntos de estudios homogéneos (p. ej., los que realicen una misma comparación).

Además de en caso de heterogeneidad, un metaanálisis tampoco es adecuado en caso de que los estudios sean de mala calidad o en presencia de un sesgo de publicación evidente, ya que podría conducir a conclusiones e interpretaciones erróneas.

Pueden diferenciarse dos modelos estadísticos básicos para combinar los resultados: el de *efectos fijos*, que sólo considera la variabilidad presente en cada estudio (intraestudio), y el modelo de *efectos aleatorios*, que considera además la variabilidad interestudios, al suponer que los estudios incluidos son una muestra de todos los posibles. Cuando no existe heterogeneidad, ambos modelos proporcionan resultados prácticamente idénticos. Cuando existe heterogeneidad, lo más adecuado es no combinar los resultados, pero si se decide hacerlo, debe utilizarse el modelo de efectos aleatorios, ya que es más conservador y proporciona IC más amplios para la estimación conjunta, advirtiendo claramente a los lectores de las limitaciones de este abordaje y de que deben interpretar los resultados con mucha precaución.

La combinación estadística de los diferentes estudios implica ponderar los resultados de cada uno de ellos de acuerdo con su precisión. Cuanto

más preciso es el resultado de un estudio, más estrecho es su IC, por lo que este estudio debe «pesar más» en el resultado final que los que proporcionan resultados menos precisos. Para realizar esta ponderación suele utilizarse el inverso de la variancia de la estimación o el tamaño de la muestra de los estudios. Con menos frecuencia se utiliza la puntuación recibida en el proceso de evaluación de su calidad. Las pruebas estadísticas más utilizadas se presentan en la tabla A6.1.

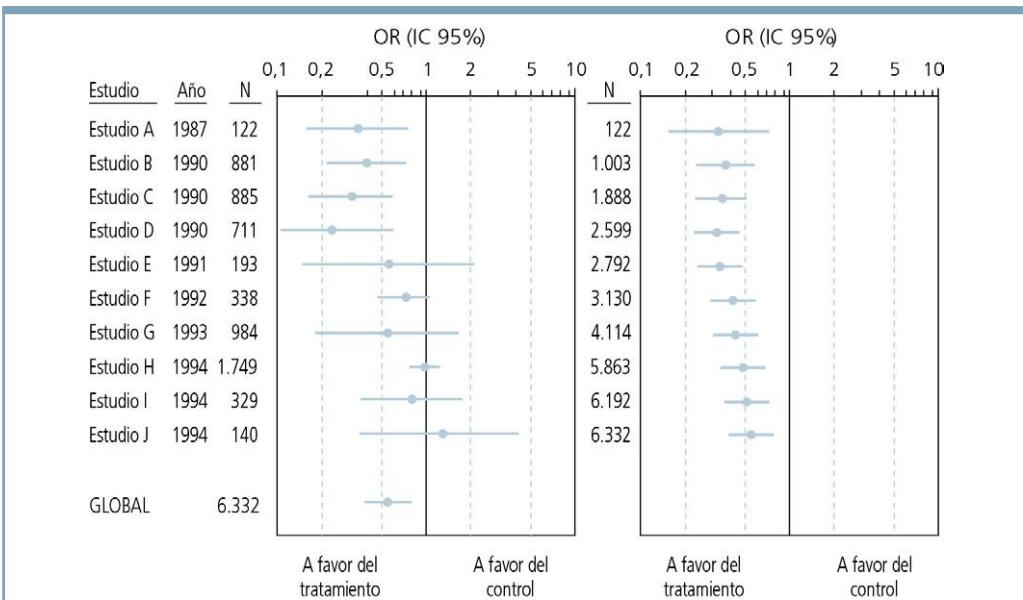
Es importante examinar si el resultado global es consistente, es decir, si está muy influido por las decisiones tomadas durante el proceso de revisión, como por ejemplo, la exclusión de un determinado tipo de estudios, los criterios de selección empleados, la utilización de un método de efectos aleatorios o fijos para combinar los datos, etc. Con esta finalidad pueden realizarse análisis de sensibilidad, metaanálisis acumulativos, análisis de subgrupos o una metarregresión.

### PRESENTACIÓN GRÁFICA DE LOS RESULTADOS

Habitualmente, los resultados de una RS suelen presentarse gráficamente en un formato relativamente estándar (fig. A6.1). El resultado principal de cada uno de los estudios incluidos en la revisión se muestra como una línea horizontal que

**Tabla A6.1 Principales pruebas estadísticas utilizadas en un metaanálisis**

Tipo de variable	Medida del efecto	Modelo	Método
Dicotómica	Odds ratio	Efectos fijos	Peto Mantel-Haenszel Woolf
	Riesgo relativo	Efectos aleatorios	DerSimonian-Laird
		Efectos fijos	Mantel-Haenszel Woolf
	Diferencia de riesgos	Efectos aleatorios	DerSimonian-Laird
		Efectos fijos	Mantel-Haenszel Woolf
		Efectos aleatorios	DerSimonian-Laird
Continua	Diferencia de medias	Efectos fijos	Woolf
		Efectos aleatorios	DerSimonian-Laird
	Diferencia estandarizada de medias	Efectos fijos	Woolf
		Efectos aleatorios	DerSimonian-Laird



**Figura A6.1** Presentación gráfica de los resultados de un metaanálisis sobre la eficacia de una intervención sanitaria. Se representan los valores de *odds ratio* de cada estudio individual y la global del metaanálisis, y sus correspondientes intervalos de confianza del 95%.

representa el IC del 95% de la estimación de la medida del efecto, y en la que una señal indica la estimación puntual observada en dicho estudio. Una última línea horizontal representa el resultado global del metaanálisis, cuya estimación puntual suele representarse por un rombo. Una línea vertical representa la ausencia de efecto (valor uno, cuando se utilizan medidas relativas como la *odds ratio* o el riesgo relativo, o valor cero, cuando se utilizan medidas absolutas como la diferencia de medias o de incidencias).

Cuando los estudios se presentan ordenados por algún criterio, como el año de su realización, se denominan *metaanálisis acumulativos* (parte derecha de la fig. A6.1), que muestra la ejecución repetida del metaanálisis cuando un nuevo estudio está disponible para su inclusión. Ayuda a determinar la consistencia de los resultados a lo largo del tiempo e identificar retrospectivamente el punto en el tiempo en que la estimación conjunta del efecto alcanzó un determinado grado de significación. La primera línea coincide en ambas figuras, mientras que la segunda línea de la figura de la derecha corresponde al análisis acumulado de los dos primeros estudios; la tercera, de tres estudios, y así, sucesivamente. La amplitud del IC se va reduciendo, ya que cada vez se analiza un mayor número de sujetos y la estimación es más precisa.

Además, puede evaluarse si la incorporación de nuevos estudios produce alguna modificación en la estimación del efecto. En la figura puede apreciarse cómo, a medida que se añaden nuevos estudios, existe una ligera tendencia a la disminución de la magnitud del efecto (se va acercando a la línea de ausencia de efecto).

## ANÁLISIS DE SENSIBILIDAD

Este procedimiento requiere reanalizar los datos en función de diferentes criterios y comprobar si se modifican los resultados: analizando solamente estudios experimentales y añadiendo observacionales, incluyendo estudios de menor calidad metodológica o con un tamaño de la muestra más reducido, etc. Si la utilización de diferentes criterios conduce a resultados similares, la conclusión de la RS se ve fortalecida.

Una forma de análisis de sensibilidad es efectuar el metaanálisis empleando un método de efectos aleatorios y un método de efectos fijos. La estimación puntual con ambos métodos será muy similar, pero el IC será más amplio con el método de efectos aleatorios. Si el IC varía muy poco, se tiene una mayor confianza en que los resultados de la RS son consistentes.

Otras características de la calidad de los estudios que pueden servir para efectuar un análisis de sensibilidad son la forma en que se asignaron los sujetos al grupo de intervención o al de control, o cómo se evaluó la variable de respuesta. Si los resultados de la RS son diferentes al incluir o excluir estudios con menor calidad metodológica, los resultados deben interpretarse con precaución.

Algo similar sucede con los estudios no publicados, de los que se ha podido obtener información. Lo más correcto, si cumplen criterios de inclusión, es incluirlos en el análisis. De todos modos, conviene analizar también la sensibilidad de los resultados globales de la RS cuando se excluyen estos estudios, lo que puede orientar sobre la posible existencia de un sesgo de publicación. Si los resultados difieren debe informarse a los lectores del hecho e interpretar los resultados con cautela.

El metaanálisis acumulativo, comentado anteriormente, corresponde a un análisis de sensibilidad que permite evaluar cómo influye el criterio por el que se ordenan los estudios, ya sea el año de publicación o la puntuación obtenida al evaluar su calidad metodológica (p. ej., empezando por los de mayor calidad y evaluando cómo la inclusión progresiva de estudios de menor calidad influye sobre la estimación combinada del efecto).

### METARREGRESIÓN

Se utiliza para evaluar estadísticamente qué factores influyen sobre la estimación del efecto, como determinadas características de los pacientes, de la intervención o del diseño de los estudios. Estas características pueden ser comunes a todos los pacientes incluidos en los estudios, como por ejemplo, la vía de administración de un tratamiento, o un promedio de todos los pacientes, como la media de edad. Este último tipo de covariables requiere una interpretación cautelosa, ya que los valores agregados pueden no representar adecuadamente a grupos minoritarios de pacientes. Los resultados de la metarregresión se informan generalmente como coeficientes de regresión con sus IC, que reflejan la influencia de cada variable evaluada sobre la estimación del efecto.

### ANÁLISIS DE SUBGRUPOS

Los análisis de subgrupos se realizan cuando interesa evaluar los resultados en algún grupo especial de sujetos, en función de la edad, el sexo u otras

variables, y solamente se justifican si se han planeado antes de iniciar la RS, a partir de una hipótesis fundamentada.

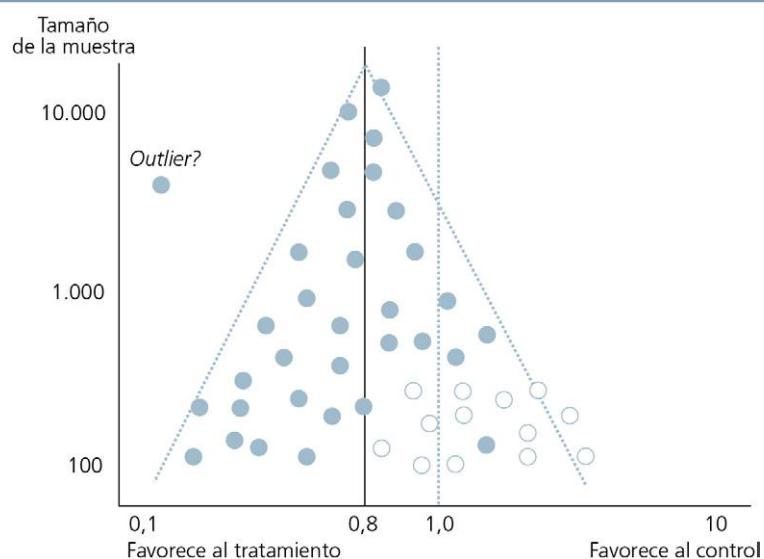
Además de los problemas de cualquier análisis de subgrupos, en una RS es importante tener en cuenta si la comparación se realiza entre subgrupos de sujetos que proceden de un mismo estudio o bien entre subgrupos procedentes de estudios diferentes. En esta segunda situación, el resultado tiene mayores limitaciones, ya que, al proceder de estudios diferentes, pueden existir otros muchos factores que influyan sobre los resultados.

Cuando la finalidad no es tanto obtener estimaciones por subgrupos como ver si los resultados difieren en función de los subgrupos que se consideren, se utiliza como otra forma de explorar y explicar la heterogeneidad.

### EVALUACIÓN GRÁFICA DEL SESGO DE PUBLICACIÓN

Una forma de evaluar la posibilidad de un sesgo de publicación es el gráfico en embudo (*funnel plot*), en el que las medidas del efecto observadas en los diferentes estudios se presentan en una gráfica cuyo eje de abscisas corresponde a la magnitud del efecto y el de ordenadas a alguna medida de su variabilidad, generalmente el tamaño muestral (fig. A6.2). Dado que los estudios de menor tamaño tienen una menor precisión en la estimación del efecto, presentan una mayor variabilidad en los resultados. En cambio, los estudios de mayor tamaño muestran mayor precisión, por lo que sus resultados tienen menor variabilidad. Al representar los resultados de los diferentes estudios identificados, se obtiene un gráfico en forma de embudo invertido o pirámide, si no existe sesgo de publicación. Cuando, como ocurre en la figura, el gráfico no es simétrico (obsérvese que parecen faltar estudios en el ángulo inferior derecho), puede sospecharse que no se han localizado los estudios que deberían ocupar este lugar; en este caso, estudios de pequeño tamaño con resultados en contra del tratamiento, lo que sería sugestivo de la existencia de un sesgo de publicación. Éste es un método aproximado y difícil de interpretar cuando existen pocos estudios.

Otra utilidad del gráfico en embudo es que permite identificar algún estudio con comportamiento extraño, como el indicado como *outlier* en la figura A6.2. Se trata de un estudio que conduce a una estimación muy favorable al tratamiento, pero que queda fuera



**Figura A6.2** Gráfico en embudo para la evaluación del sesgo de publicación.

de la variabilidad que sería esperable simplemente por azar en los estudios de un tamaño de la muestra similar. Ello obliga a revisar cuidadosamente

las características de este estudio con la finalidad de identificar algún factor que pueda explicar este comportamiento inesperado.

## BIBLIOGRAFÍA

- Egger M, Smith GD, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. 2nd ed. London: BMJ Publishing Group; 2001.
- Friedrich JO, Adhikari NK, Beyene J. Ratio of means for analyzing continuous outcomes in meta-analysis performed as well as mean difference methods. *J Clin Epidemiol*. 2011;64:556-64.
- Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analysis. *Int J Epidemiol*. 2002;31:72-6.
- Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Disponible en: [www.cochrane-handbook.org](http://www.cochrane-handbook.org)
- Higgins J, Thompson S, Deeks JJ, Altman D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *J Health Serv Res Policy*. 2002;7:51-61.
- Lau J, Ioannidis JPA, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med*. 1997;127:820-6.
- Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Stat Med*. 2001;20:641-54.
- Palma Pérez S, Delgado Rodríguez M. Consideraciones prácticas acerca de la detección del sesgo de publicación. *Gac Sanit*. 2006;20(Supl3):10-6.
- Pereira TV, Ioannidis JP. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *J Clin Epidemiol*. 2011;64:1060-9.
- Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ*. 2011;342: d549. doi: 10.1136/bmj.d549.
- Sedgwick P. Meta-analyses: funnel plots. *BMJ*. 2011;343:d5372.
- Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol*. 2001;54:1046-55.
- Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analysis. *BMJ*. 2000;320:1574-7.

En este anexo se presentan, a modo de enunciados o preguntas, los aspectos más importantes que hay que tener en cuenta cuando se elabora un protocolo de estudio. Los ítems no deben considerarse estrictamente en el orden en que se presentan, ya que son un simple recordatorio de los puntos sobre los que hay que reflexionar a medida que se diseña un trabajo.

### INTRODUCCIÓN: ANTECEDENTES Y JUSTIFICACIÓN

1. *Formule la pregunta de interés como una frase interrogativa sencilla. A medida que desarrolle el proyecto, encontrará útil reescribir esta pregunta varias veces, formulándola cada vez con más precisión.*
2. *Identifique las palabras clave relacionadas con la pregunta. Siempre que sea posible, utilice términos empleados en la lista del Medical Subject Headings (MeSH) de Medline.*
3. *Realice una búsqueda bibliográfica utilizando las palabras clave. Identifique las citas más relevantes. Consulte con expertos en el tema. Identifique las teorías relevantes y otras fuentes de información destacadas. Elabore una ficha de las citas principales, anotando sus aspectos más importantes y los motivos por los que un artículo se considera de interés para esta investigación.*
4. *Resuma la situación actual del conocimiento sobre el tema del estudio.*
5. *Formule la hipótesis de trabajo.*
6. *Valore la factibilidad de la investigación. Evalúe si podrá medir de forma precisa y exacta las variables de interés, el tiempo necesario y la disponibilidad de sujetos, la posibilidad de colaboración de otros profesionales o centros, las instalaciones, los equipos y los recursos disponibles, y si el equipo investigador tiene suficiente experiencia.*
7. *Valore la pertinencia del estudio. ¿Qué importancia puede tener la respuesta correcta a la pregunta? ¿Qué implicaciones pueden tener el resto de respuestas posibles? Evalúe si los beneficios potenciales compensan los recursos necesarios para realizar el estudio.*
8. *Justifique la realización del estudio.*

### OBJETIVOS ESPECÍFICOS

9. *Formule el objetivo principal del estudio en forma operativa, especificando el factor de estudio, la variable de respuesta principal y la población que*

*se desea estudiar. Especifique si desea evaluar superioridad, equivalencia o no inferioridad.*

10. *Formule otros objetivos secundarios o adicionales.*

### APLICABILIDAD Y UTILIDAD DE LOS RESULTADOS

11. *Identifique los beneficios potenciales derivados de la respuesta. ¿Qué aplicabilidad tienen? ¿A quién afectan?*

### DISEÑO Y MÉTODOS

12. *Describa en una frase el tipo de diseño que utilizará.*
13. *Describa la población a la que desearía generalizar los resultados (población diana).*
14. *Describa las características de las personas que desearía estudiar (población de estudio). Defina con claridad y precisión los criterios de selección (de inclusión y de exclusión) de los sujetos. Valore definiciones estándar utilizadas por otros investigadores. Contraste la validez interna con la capacidad de generalización.*
15. *Determine el tamaño de la muestra necesario, especificando las asunciones utilizadas para el cálculo. Defina la mínima diferencia o asociación de relevancia clínica que desea ser capaz de detectar.*
16. *Describa cuál será la procedencia de los sujetos, cómo se identificarán los candidatos y cómo se seleccionará la muestra. Si utiliza alguna técnica de muestreo, especifique cuál.*
17. *Decida si dividirá la muestra en subgrupos. En caso afirmativo, ¿cómo lo hará? Si desea realizar un ensayo clínico, describa el tipo de asignación y cómo se realizará.*
18. *Si se utiliza alguna técnica de enmascaramiento, descríbala.*
19. *Elabore el marco teórico en que se sitúa el estudio y confeccione una lista de todas las variables relevantes que deberían ser medidas. Identifique las relaciones de cada una de ellas con el factor de estudio y la variable de respuesta, así como su posible papel en la relación entre ellos.*
20. *Para cada una de las variables, elabore una definición operativa e identifique el instrumento de medida o fuente de información de que dispone. En la medida de lo posible, utilice definiciones e instrumentos estándar para la*

- medición de las variables. Elabore definiciones claras y precisas para que todos los investigadores utilicen criterios homogéneos. Identifique las fuentes de información necesarias.
21. *Si no dispone de un instrumento adecuado para medir alguna variable relevante, identifique las características necesarias que le permitan buscarlo o desarrollarlo.*
  22. *Para cada instrumento de medida, valore la fiabilidad y la validez, identificando los potenciales problemas. Establezca las estrategias necesarias para mejorarlas. Valore la posibilidad y la conveniencia de utilizar técnicas tales como la selección de medidas objetivas antes que subjetivas, la formación de los observadores, la utilización de instrumentos más precisos, la obtención de varias mediciones de una variable o el empleo de técnicas de enmascaramiento.*
  23. *Identifique los potenciales factores de confusión que deberán controlarse y aquellas variables que puedan modificar el efecto del factor de estudio sobre la respuesta.*
  24. *Si realiza alguna intervención, descríbala detalladamente.*
  25. *Describa lo que le ocurrirá a cada persona que participe en el estudio: cómo se comprobará si cumple los criterios de selección, cómo se solicitará su consentimiento, qué seguimiento se hará, cómo recibirá la intervención, qué mediciones se realizarán, qué se hará si se presenta cualquier acontecimiento durante el estudio, etc.*
  26. *Describa quién recogerá los datos y cómo lo hará. Establezca los circuitos necesarios para la derivación a exploraciones complementarias y el procesamiento de muestras, así como para el registro de sus resultados.*
  27. *Haga un esquema de los impresos necesarios para la recogida de datos. Diseñe los impresos de forma que faciliten la recogida de datos y minimicen los posibles errores en este proceso.*
  28. *Establezca los mecanismos necesarios para el control de la calidad de los datos recogidos.*
  29. *Describa quién realizará el procesamiento informático de los datos, cómo y cuándo se hará.*
  31. ¿Es posible que las personas que se pierdan durante el estudio sean diferentes de las que permanezcan en él? En caso afirmativo, describa el problema y su posible solución. Diseñe estrategias para minimizar el número de pérdidas durante el seguimiento y para recaptarlas si se producen.
  32. ¿Es posible que el instrumental se deteriore o los observadores se aburran? En caso afirmativo, describa el problema y su posible solución. Diseñe estrategias que aseguren que la calidad y la homogeneidad de los datos se mantienen a lo largo de todo el estudio. Establezca mecanismos de monitorización que le permitan identificar fluctuaciones en los datos que sugieran alguno de estos fenómenos.
  33. ¿Puede anticipar si es posible que durante el estudio ocurran cambios de personal u otros sucesos que interfieran en los resultados? En caso afirmativo, describa el problema y su posible solución.
  34. ¿Podría usted, como investigador, influir sobre los resultados? En caso afirmativo, describa el problema y su posible solución.
  35. Identifique otras posibles fuentes de sesgo en su estudio y sus posibles soluciones.
  36. Identifique otras limitaciones del diseño del estudio que deban tenerse en cuenta al interpretar los resultados.
  37. Asegure que el diseño del estudio cumple los requisitos éticos.

## ESTRATEGIA DE ANÁLISIS

### LIMITACIONES Y POSIBLES SESGOS DEL ESTUDIO

30. ¿Hay algo en la selección de sujetos o en su distribución en subgrupos que pueda hacer que los grupos no sean comparables? En caso afirmativo, describa el problema y su posible solución.

38. Identifique las variables demográficas que utilizará para describir la población estudiada.
39. Identifique las principales variables pronósticas que utilizará para valorar la comparabilidad de los grupos.
40. Defina la medida que utilizará para cuantificar el efecto o la asociación de interés. Identifique las pruebas estadísticas que necesitará aplicar.
41. Defina las técnicas estadísticas que utilizará para ajustar o controlar el efecto de los posibles factores de confusión, si es necesario.
42. Defina la estrategia de análisis, especificando qué sujetos incluirá, cómo tratará las no respuestas, pérdidas e incumplimientos del protocolo y cómo valorará su posible influencia sobre los resultados.

**43.** Identifique los subgrupos de sujetos en los que querrá realizar el análisis, y justifique los motivos.

### ORGANIZACIÓN DEL ESTUDIO

- 44.** Prepare todos los documentos necesarios para llevar a cabo el estudio: formularios para la recogida de datos, cartas de presentación, formularios de consentimiento informado, solicitudes de colaboración, etc.
- 45.** Identifique las tareas administrativas a realizar y las personas que las llevarán a cabo.
- 46.** Identifique las personas y los servicios con los que deberá contactar. Asegure su colaboración y

describa los circuitos adecuados para llevarla a cabo.

**47.** Elabore un cronograma con todas las fases del estudio. Identifique las responsabilidades de los miembros del equipo investigador. Fije un calendario de reuniones.

**48.** Realice una prueba piloto.

**49.** Someta el protocolo a un comité ético para su valoración. Compruebe que se cumplen los requisitos éticos y legales aplicables al estudio.

### PRESUPUESTO

**50.** Realice una estimación del presupuesto necesario. Asegure la financiación antes del inicio del estudio.

### EVALUACIÓN DE PRUEBAS DIAGNÓSTICAS

El análisis de este tipo de estudios se basa fundamentalmente en el cálculo de la sensibilidad y la especificidad de la prueba diagnóstica, que son dos proporciones. Por consiguiente, el cálculo del tamaño de la muestra corresponde a la situación de estimación de una proporción.

**Ejemplo A8.1** Supongamos que un investigador desea determinar la sensibilidad de una nueva prueba diagnóstica. Basándose en un estudio piloto, espera que el 85% de los pacientes tenga resultados positivos con esta prueba. Desea estimar este valor con una precisión del 5% y un nivel de confianza del 95%. Aplicando la fórmula del cuadro 15.1 (estimación de un parámetro; v. pág. 143), se calcula que se requieren 196 sujetos con la enfermedad.

Si el estudio se diseña para seleccionar una sola muestra de individuos, debe conocerse la prevalencia de enfermedad en la población que cumple los criterios de selección, para determinar cuántos sujetos deberán incluirse en el estudio para que existan los 196 enfermos necesarios. Por ejemplo, si se supone que entre los sujetos que van a ser estudiados, el 40% tendrá la enfermedad, será necesario iniciar el estudio con una muestra de 490 individuos.

Si el estudio se plantea de forma que se deba seleccionar una muestra de pacientes y otra de controles, el número de estos últimos se calculará de la misma forma que el de los casos, pero enfocando la situación como la estimación de la especificidad esperada.

**Ejemplo A8.2** El investigador del ejemplo A8.1 desea determinar la especificidad de la prueba para descartar la enfermedad, y espera que el 90% de los individuos sin la enfermedad tengan la prueba negativa. Desea realizar esta estimación con la precisión del 5% y la confianza del 95%. Utilizando la misma fórmula, determinará que son necesarios 138 sujetos sin la enfermedad.

### EVALUACIÓN DE LA CONCORDANCIA

Cuando el objetivo del estudio es estimar la concordancia entre los observadores en el caso de una variable cualitativa, los datos se analizan

mediante el índice Kappa. El cálculo del tamaño de la muestra necesario para estimar este índice con una precisión determinada conlleva problemas prácticos, ya que obliga a prefijar no sólo el porcentaje total de discordancia esperado, sino también el esperado por azar. Debido a que el cálculo del tamaño de la muestra depende de tantas suposiciones, el resultado podría no ser el más adecuado.

Desde un punto de vista práctico, se asume que el parámetro que se estima es el porcentaje de desacuerdo entre observadores. Se trata, por tanto, de la estimación de una proporción.

**Ejemplo A8.3** Supongamos que se desea evaluar el grado de concordancia entre dos radiólogos en la lectura de mamografías. En un estudio previo se ha demostrado que presentan una elevada repetibilidad, por lo que sólo interesa evaluar la concordancia entre observadores. Se espera que el porcentaje de desacuerdo entre ellos sea de un 20%, aproximadamente. Interesa estimar dicho porcentaje con una precisión del 5% y una confianza del 95%. Aplicando la fórmula del cuadro 15.1 (estimación de un parámetro; v. pág. 143), se calcula que ambos radiólogos deberán leer 246 mamografías.

Cuando el objetivo es estimar la concordancia intraobservador, los supuestos y el cálculo del tamaño de la muestra necesario son los mismos que para analizar la repetibilidad de una medida. El cálculo que se presenta corresponde a la situación en que cada observador realizará dos mediciones de cada sujeto y se desea evaluar el grado de concordancia de ambas mediciones. Lo primero que debe determinarse es el porcentaje de error (discordancia) entre ambas medidas que se espera encontrar, así como la precisión con que interesa realizar la estimación y la confianza con que se quiere expresar el resultado. En la tabla H del final del libro se presentan los resultados de este cálculo para las situaciones más habituales.

**Ejemplo A8.4** Consideremos que se desea estimar la concordancia entre dos lecturas de mamografías realizadas por un mismo radiólogo. Se espera que la discordancia intraobservador se sitúe alrededor del 5% ( $e = 0,05$ ). Se requiere una precisión del 5% ( $i = 0,05$ ) y una confianza del 95% ( $Z_{\alpha} = 1,96$ ). Utilizando la tabla H, se determina que es necesario que el radiólogo lea, en dos ocasiones, 41 mamografías, aproximadamente.

En muchas ocasiones, se diseña el estudio para evaluar simultáneamente la concordancia intraobservador e interobservador. La tabla I del final del libro sirve para determinar el número de sujetos necesarios cuando existen dos observadores que evalúan cada uno de ellos en dos ocasiones a todos los sujetos de la muestra.

**Ejemplo A8.5** En un estudio hipotético se desea evaluar simultáneamente la concordancia intraobservador e interobservador de dos radiólogos en la lectura de mamografías. Se asume que el porcentaje de discordancia entre ambos radiólogos se situará alrededor del 0,15, y que la probabilidad de error de cada radiólogo es de 0,05. Interesa que el intervalo de confianza del 95% tenga una amplitud total de 0,10 ( $2i = 0,10$ ). Serán necesarias 252 mamografías, cada una de las cuales será leída en cuatro ocasiones: dos por cada radiólogo.

### ESTIMACIÓN DE UN COEFICIENTE DE CORRELACIÓN

La asociación entre dos variables cuantitativas requiere habitualmente la utilización del coeficiente de correlación  $r$  de Pearson. El cuadro A8.1 presenta la fórmula para el cálculo del número de sujetos necesarios en esta situación. La tabla J del final del libro simplifica esta determinación.

Cuadro A8.1 Fórmula para el cálculo del número de sujetos necesarios en un estudio cuyo objetivo es evaluar una asociación entre dos variables cuantitativas, utilizando el coeficiente de correlación de Pearson

$$N = \left\lceil \frac{(Z_\alpha + Z_\beta)^2}{0,5 \ln[(1+r)/(1-r)]} \right\rceil + 3$$

N: número de sujetos necesarios.

$Z_\alpha$ : valor de Z correspondiente al riesgo  $\alpha$  fijado (tabla 15.4).

$Z_\beta$ : valor de Z correspondiente al riesgo  $\beta$  fijado (tabla 15.5).

In: logaritmo natural o neperiano.

r: valor del coeficiente de correlación que se supone existe en la población.

**Ejemplo A8.6** El objetivo de un estudio es determinar si existe asociación entre la presión arterial sistólica y la edad. Según la literatura científica, puede esperarse un valor del coeficiente de correlación cercano a 0,7. A continuación, se fijan los niveles de error ( $\alpha = 0,05$ ;  $\beta = 0,10$ ). Utilizando la tabla J se obtiene un tamaño de 17 individuos.

### ENSAYOS CRUZADOS

Los diseños cruzados precisan un número menor de sujetos, ya que cada uno de ellos actúa como su propio control, disminuyendo la variabilidad de las respuestas. En este tipo de diseños, la variancia tiene dos componentes. Por un lado, la variabilidad intersujetos de la respuesta, y por otro, la variabilidad debida al azar y a otros factores intrasujetos. El cálculo requiere incorporar esta descomposición de la variancia, para lo cual puede realizarse como si se tratara de un estudio de grupos paralelos, ajustando posteriormente el número obtenido en función del *coeficiente de correlación intraclasa*, que indica la proporción de la variabilidad total que se debe al componente intersujetos. Incluso si este coeficiente es tan sólo de 0,5, es decir, que sólo la mitad de la variabilidad se debe a factores intersujetos, el tamaño de la muestra puede reducirse a una cuarta parte del necesario en un estudio paralelo. Dado que este coeficiente es de difícil cálculo, una norma segura es utilizar la mitad del tamaño necesario para un estudio paralelo, que asegura una potencia

Cuadro A8.2 Fórmula para el cálculo del tamaño de la muestra en un ensayo cruzado (variable cuantitativa)

$$N = \frac{(Z_\alpha + Z_\beta)^2 \cdot S_d^2}{d^2}$$

N: número de sujetos necesarios.

$Z_\alpha$ : valor de Z correspondiente al riesgo  $\alpha$  fijado (tabla 15.4).

$Z_\beta$ : valor de Z correspondiente al riesgo  $\beta$  fijado (tabla 15.5).

$S_d^2$ : variancia muestral de las diferencias individuales.

d: valor mínimo de la diferencia que se desea detectar.

estadística suficiente. De todas formas, una fórmula aproximada se presenta en el cuadro A8.2.

**Ejemplo A8.7** Supongamos que se diseña un ensayo cruzado para comparar dos tratamientos hipoglucemiantes. La diferencia mínima que se desea detectar es de 3 mg/dl. De estudios anteriores se sabe que la variancia muestral de las diferencias es 290. Se acepta un riesgo  $\alpha$  bilateral de 0,05 y una potencia estadística del 90% ( $\beta = 0,10$ ). Aplicando la fórmula, se obtiene que son necesarios 338 pacientes.

detectar un riesgo como mínimo dos veces superior en la cohorte expuesta en relación con la no expuesta ( $\delta = 2$ ), aceptando un error  $\alpha$  bilateral de 0,05 y  $\beta$  de 0,20 (potencia:  $1 - \beta = 0,80$ ). La tabla K indica que serán necesarios 33 desenlaces por grupo. Este número indica el número de desenlaces que deberán observarse, no el de sujetos que deberán iniciar el estudio. Para calcular el número de personas que deberán iniciar el estudio, es preciso conocer o asumir el porcentaje de personas que se espera que presenten el desenlace en la cohorte no expuesta. Si se espera que el 10% de los sujetos de la cohorte de referencia desarrollará el desenlace, deberán incluirse  $33/0,10 = 330$  individuos.

## ESTUDIOS DE SUPERVIVENCIA

En los estudios en que el período de seguimiento no es el mismo en todos los sujetos, ya sea porque entran en el estudio en fechas diferentes o porque se pierden durante el seguimiento, se utilizan técnicas de análisis que tienen en cuenta este hecho (análisis de supervivencia).

Para calcular el tamaño de la muestra en estas situaciones, es necesario fijar la magnitud del efecto que interesa detectar, que corresponde al cociente entre el riesgo de desarrollar el suceso en un momento determinado del tiempo en la cohorte expuesta y en la cohorte no expuesta. A esta razón se la denomina  $\delta$ , y es un concepto análogo al de riesgo relativo. El número de desenlaces  $\delta$  en cada grupo se puede determinar con la siguiente fórmula:

$$\delta = [2(Z_{\alpha} + Z_{\beta})^2] / \ln\delta^2$$

Donde  $\ln \delta$  indica el logaritmo neperiano de  $\delta$ .

La tabla K del final del libro presenta el resultado del cálculo en las situaciones más habituales.

**Ejemplo A8.8** Se quiere realizar un estudio de cohortes en el que existan diferentes tiempos de seguimiento para cada sujeto, y en el que interesa

## ESTUDIOS DE CASOS Y CONTROLES CON MÁS DE UN CONTROL POR CASO

En los estudios de casos y controles, a veces puede ser interesante utilizar dos o incluso tres controles por cada caso. Una forma sencilla para el cálculo en esta situación es:

$$m = [r + 1]n/2r$$

Donde  $n$  es el número de sujetos necesario si los grupos fueran de igual tamaño;  $m$  corresponde al número de casos, y  $r$  al número de controles por caso, por lo que  $r \cdot m$  indica el número de controles necesarios.

**Ejemplo A8.9** En el ejemplo 15.6 se había calculado que para realizar un estudio de casos y controles sobre la asociación entre el infarto agudo de miocardio y el consumo de anticonceptivos orales eran necesarios 74 casos y 74 controles. Si se decide utilizar 2 controles por caso, aplicando la fórmula se determina que  $m = 55,5$ , por lo que serán necesarios 56 casos y 112 controles.

Este anexo pretende proporcionar unas normas sencillas que faciliten la elección de la prueba estadística más adecuada para el análisis de los datos. Los métodos para su cálculo y las asunciones en que se basan están disponibles en múltiples textos de estadística.

### PRUEBAS ESTADÍSTICAS BIVARIANTES

La elección de la prueba estadística depende de (cuadro A9.1):

- La *escala de medida de la variable de respuesta*. Las pruebas estadísticas tienen mayor potencia si la variable de respuesta es cuantitativa, ya que contiene más información que si es cualitativa.
- La *escala de medida del factor de estudio*. Puede ser cualitativa dicotómica (tratamiento activo/placebo, exposición/no exposición), cualitativa con más de dos categorías (tres pautas terapéuticas o diferentes grados de exposición a un factor de riesgo) o cuantitativa (valores de la colesterolemia o la presión arterial).
- El carácter *apareado o independiente* de los datos. Desde el punto de vista estadístico, se habla de medidas repetidas o apareadas cuando han sido realizadas sobre los mismos sujetos (p. ej., comparación de las cifras de presión arterial obtenidas en los individuos de una muestra al inicio y al final de un determinado período). Dado que los sujetos son los mismos, existe una menor variabilidad en las mediciones, lo que permite utilizar pruebas más potentes que tengan en cuenta este fenómeno. En caso de que los grupos que se comparan estén formados por individuos diferentes, se habla de datos independientes.
- Las *condiciones de aplicación específicas* de cada prueba. Las pruebas estadísticas que utilizan datos cuantitativos suelen realizar determinadas asunciones sobre la distribución de las variables en las poblaciones que están siendo comparadas. Estas pruebas son conocidas como *pruebas paramétricas*. La mayoría son robustas, es decir, toleran relativamente violaciones de estas asunciones, especialmente si el número de sujetos estudiado es elevado. En muchas situaciones, especialmente cuando las muestras son de pequeño tamaño, no se puede determinar si se cumplen dichas asunciones. En estos

casos, se recurre a otras pruebas estadísticas menos potentes, que no requieren asunciones para su aplicabilidad, conocidas como *pruebas no paramétricas*. Este mismo tipo de pruebas es aplicable cuando se trata de analizar datos ordinales.

En la [tabla A9.1](#) se resumen las pruebas estadísticas que se utilizan en las situaciones más frecuentes.

De forma esquemática, cuando tanto el factor de estudio como la variable de respuesta son variables cualitativas, la prueba estadística más apropiada para determinar si existe asociación entre ellas es la  $\chi^2$  al cuadrado, siempre que exista un número suficiente de sujetos en cada una de las casillas de la tabla de contingencia.

**Ejemplo A9.1** Un estudio hipotético pretende comparar el porcentaje de diabéticos controlados en un grupo que ha recibido educación sanitaria grupal con el de otro grupo que ha recibido los cuidados habituales. Se trata de datos independientes, ya que son grupos de sujetos diferentes. El factor de estudio (educación sanitaria) es una variable dicotómica (grupal/habitual), al igual que la variable de respuesta (controlado/no controlado). Por lo tanto, los datos se presentarán en una tabla  $2 \times 2$  y se utilizará una  $\chi^2$  al cuadrado (o una Z de comparación de dos proporciones), siempre que el número de sujetos lo permita. Si no puede utilizarse, se recurrirá a la prueba exacta de Fisher.

Cuando se comparan dos grupos (factor de estudio dicotómico) respecto a una variable cuantitativa (comparación de dos medias), la prueba estadística más adecuada es la t de Student-Fisher. Si no se cumplen las condiciones necesarias para su aplicación, se recurre a una prueba no paramétrica equivalente, como la U de Mann-Whitney.

**Ejemplo A9.2** Supongamos que en el ejemplo A9.1 interesa comparar la glucemia de ambos grupos de sujetos al finalizar el estudio. En este caso, la variable de respuesta es cuantitativa (glucemia). Se trata, por tanto, de la comparación de dos medias en grupos independientes, y la prueba de elección es la t de Student-Fisher, si se cumplen las condiciones de aplicación.

Si se comparan más de dos grupos (factor de estudio con más de dos categorías) respecto a una

**Cuadro A9.1 Factores de los que depende la elección de la prueba estadística más adecuada**

- Escala de medida de la variable de respuesta
- Escala de medida del factor de estudio
- Carácter apareado o independiente de los datos
- Condiciones de aplicación específicas de cada prueba:
  - Número de observaciones
  - Asunciones sobre la distribución poblacional de las variables

su aplicación, se recurre a la prueba de Kruskal-Wallis.

**Ejemplo A9.3** Supongamos que en el ejemplo A9.1 se han incluido pacientes procedentes de tres centros de salud, y que interesa determinar si la glucemia de los sujetos difiere según el centro de procedencia. En este caso, el factor de estudio (centro de salud de procedencia) tiene tres categorías, por lo que no puede aplicarse la t de Student-Fisher. Si se cumplen las condiciones para su utilización, se aplicará un análisis de la variancia. Si se observa una diferencia estadísticamente significativa, se

**Tabla A9.1 Pruebas bivariantes de significación estadística utilizadas con mayor frecuencia**

Factor de estudio	Variable de respuesta			
	Cualitativa nominal (dos categorías)	Cualitativa nominal (> 2 categorías)	Cualitativa ordinal	Cuantitativa*
<b>Cualitativo (dos grupos)</b>				
Independientes	Z de comparación de proporciones Ji al cuadrado Prueba exacta de Fisher	Ji al cuadrado	U de Mann-Whitney	t de Student-Fisher
Apareados	Prueba de McNemar	Q de Cochran	Prueba de los signos	t de Student-Fisher para datos apareados
	Prueba exacta de Fisher		Prueba de los rangos signados de Wilcoxon	
<b>Cualitativo (más de dos grupos)</b>				
Independientes	Ji al cuadrado	Ji al cuadrado	Prueba de Kruskal-Wallis	Análisis de la variancia
Apareados	Q de Cochran	Q de Cochran	Prueba de Friedman	Análisis de la variancia de dos vías
Cuantitativo	t de Student-Fisher	Análisis de la variancia	Correlación de Spearman Tau de Kendall	Correlación de Pearson Regresión lineal

\*Cuando las pruebas estadísticas aplicables a las variables cuantitativas no cumplen las asunciones necesarias para su aplicación, deben utilizarse las pruebas correspondientes como si la variable de respuesta fuera una variable ordinal (pruebas no paramétricas).

variable cuantitativa (comparación de más de dos medias), debe utilizarse el análisis de la variancia (ANOVA). Si no se cumplen los criterios para

realizará un análisis posterior para determinar qué grupo o grupos difieren, utilizando alguna de las técnicas de comparación múltiple específicamente

diseñadas para ello (p. ej., la de Scheffé o la de Neuman-Keuls).

Si se trata de determinar la posible asociación entre un factor de estudio y una variable de respuesta cuantitativos, la prueba adecuada es la correlación de Pearson o, si no se cumplen las condiciones para su aplicación, la correlación no paramétrica de Spearman. En el caso de que pueda asumirse una relación de dependencia lineal de una de las variables respecto a la otra, se habla de regresión lineal simple.

**Ejemplo A9.4** Consideremos que se quiere evaluar si existe asociación entre la edad de un grupo de sujetos y sus cifras de presión arterial sistólica (PAS). Dado que ambas variables están medidas en una escala cuantitativa, se trata de un análisis de correlación. Si se desea determinar en cuánto se incrementa la PAS a medida que aumenta la edad de los sujetos, se realizaría un análisis de regresión lineal, en el que la PAS sería la variable dependiente, y la edad, la independiente. En cambio, no tendría sentido evaluar en cuánto aumenta la edad de los sujetos al incrementarse su PAS.

la variable dependiente como las independientes son cuantitativas, aunque en la práctica permite que algunas de las variables independientes sean cualitativas. El valor de un coeficiente de regresión lineal múltiple es una estimación del efecto de la variable independiente correspondiente sobre la variable dependiente, ajustado por el resto de las variables independientes de la ecuación. Es decir, representa el cambio esperado de la variable dependiente cuando se incrementa en una unidad el valor de la variable independiente, asumiendo que el resto de las variables del modelo se mantienen constantes.

**Ejemplo A9.5** En un estudio se evalúa la asociación entre la edad, el índice de masa corporal (IMC) y la clase social con la PAS. Dado que la variable dependiente (PAS) es cuantitativa, y que dos de las variables independientes también lo son (edad e IMC), se utiliza una regresión lineal múltiple. Al final del análisis se obtiene un coeficiente de regresión de la edad de 0,5, lo cual indica que la PAS es, en promedio, 0,5 mmHg más elevada por cada año más de edad de los sujetos, suponiendo que la clase social y el IMC sean constantes; es decir, el efecto de la edad está ajustado por estas dos variables.

Cuando la variable dependiente es dicotómica (enfermo/no enfermo, curado/no curado, etc.), se utiliza la regresión logística. Mientras que en la regresión lineal los coeficientes representan directamente el cambio en la variable dependiente, en la regresión logística estiman medidas relativas; por ejemplo, la *odds ratio* asociada al factor de estudio y ajustada por el resto de variables independientes. Dado que el modelo logístico no es lineal, sino exponencial, se utilizan transformaciones logarítmicas que hacen que los coeficientes no puedan interpretarse directamente.

**Ejemplo A9.6** Un estudio evalúa la relación entre diferentes factores de riesgo y el desarrollo de cardiopatía coronaria en una muestra de varones adultos. Dado que la variable de respuesta es dicotómica (desarrollo o no de la enfermedad), se realiza un análisis de regresión logística con esta variable como dependiente y el conjunto de factores de riesgo como independientes. En la ecuación resultante, la variable edad tiene un coeficiente de 0,12. Para poder interpretarlo, debe calcularse su antilogaritmo natural  $e^{0,12} = 1,13$ , valor que corresponde a la *odds ratio* asociada al aumento de

### TÉCNICAS ESTADÍSTICAS MULTIVARIANTES

En investigación clínica y epidemiológica, las técnicas multivariantes se utilizan habitualmente cuando existe una variable dependiente (variable de respuesta) y múltiples variables independientes (factor de estudio y otras variables que controlar). Estas técnicas se emplean con finalidades de *predicción* (obtener una ecuación que permita, conociendo los valores de un conjunto de variables independientes, predecir el valor de la variable dependiente), de *descripción* de la relación entre variables (identificar, de entre un conjunto de variables independientes, cuáles están asociadas con la variable dependiente), o de *estimación* del efecto del factor de estudio (obtener una estimación del efecto del factor de estudio sobre la variable de respuesta, controlando la influencia de variables de confusión). La técnica elegida dependerá de las escalas de medida de las variables implicadas en el análisis ([tabla A9.2](#)).

Una de las técnicas más utilizada es la regresión lineal múltiple, que se aplica cuando tanto

**Tabla A9.2 Elección de la técnica multivariante adecuada**

Técnicas inferenciales* (pueden diferenciarse variables dependientes e independientes)		
Variable dependiente	Variables independientes	Técnica multivariante
Cuantitativa	Cualitativas	Análisis de la variancia
	Cuantitativas (1)	Regresión lineal múltiple
Cualitativa	Cualitativas	Modelo log-lineal
	Cuantitativas (1)	Regresión logística (2) Análisis discriminante (3) Regresión de Cox (4)

Técnicas descriptivas (todas las variables presentan el mismo estatus)	
Finalidad	Técnica multivariante
Reducir la dimensionalidad de la matriz de datos (identificar los factores subyacentes en un conjunto de variables)	Análisis factorial Análisis de componentes principales Análisis de correspondencias
Clasificar a individuos o variables en grupos homogéneos	Análisis de conglomerados ( <i>clusters</i> )

\*Sólo se presentan las principales pruebas aplicables cuando existe una sola variable dependiente. (1) Algunas variables independientes pueden ser cualitativas. (2) Suele utilizarse cuando la variable dependiente es dicotómica. (3) Suele utilizarse cuando la variable dependiente tiene más de dos categorías. (4) Se utiliza cuando la variable dependiente incluye la dimensión tiempo (análisis de supervivencia).

un año de edad de los sujetos, ajustada por el resto de las variables de la ecuación.

También se utiliza con frecuencia el modelo de regresión de Cox, que es útil cuando la variable dependiente es el tiempo de supervivencia o

el tiempo transcurrido hasta la aparición de un desenlace determinado.

Existen técnicas multivariantes útiles cuando no existe una variable que pueda ser considerada dependiente, pero se utilizan poco en investigación clínica y epidemiológica.

## BIBLIOGRAFÍA

- |   |  |   |
|---|--|---|
| Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991. | Oxford: Blackwell Science; 2002.   | práctica clínica. 6. <sup>a</sup> ed. Madrid: Elsevier; 2008. |
| Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. 4th ed. | Jiménez Villa J. Estadística. En: Martín Zurro A, Cano Pérez JF, editors. Atención primaria: conceptos, organización y | Bioestadística. Madrid: Harcourt Brace de España; 1998.       |

Página deliberadamente en blanco

# Tablas

- TABLA A** Número de sujetos necesarios para la estimación de una proporción (variable cualitativa) **371**
- TABLA B** Número de sujetos necesarios para la estimación de una media (variable cuantitativa) **372**
- TABLA C** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde a la proporción menor de las dos que se comparan **374**
- TABLA D** Número de sujetos necesarios en cada grupo de estudio para estimar un riesgo relativo (RR) **385**
- TABLA E** Número de sujetos necesarios en cada grupo de estudio para estimar una *odds ratio* (OR) **388**
- TABLA F** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos medias mediante la prueba de la t de Student-Fisher **390**
- TABLA G** Número de sujetos necesarios en cada grupo en estudios de equivalencia **393**
- TABLA H** Número de sujetos necesarios para la estimación de la concordancia intraobservador **395**
- TABLA I** Número de sujetos necesarios para la estimación de la concordancia intra e interobservador **395**
- TABLA J** Número de sujetos necesarios para estimar una asociación entre dos variables cuantitativas mediante el coeficiente de correlación de Pearson **396**
- TABLA K** Número de sujetos necesarios por grupo para detectar una determinada razón de riesgos en estudios que utilizan análisis de supervivencia **397**
- TABLA L** Números aleatorios **398**

Página deliberadamente en blanco

**Tabla A. Número de sujetos necesarios para la estimación de una proporción (variable cualitativa)**

Proporción esperada (P)	Nivel de confianza (1- $\alpha$ )	Precisión (i)									
		0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,10
0,01	0,95	380	—	—	—	—	—	—	—	—	—
	0,99	657	—	—	—	—	—	—	—	—	—
0,02	0,95	753	188	—	—	—	—	—	—	—	—
	0,99	1.301	325	—	—	—	—	—	—	—	—
0,03	0,95	1.118	279	124	—	—	—	—	—	—	—
	0,99	1.931	483	215	—	—	—	—	—	—	—
0,04	0,95	1.475	369	164	92	—	—	—	—	—	—
	0,99	2.548	637	283	159	—	—	—	—	—	—
0,05	0,95	1.825	456	203	114	73	—	—	—	—	—
	0,99	3.152	788	350	197	126	—	—	—	—	—
0,06	0,95	2.167	542	241	135	87	60	—	—	—	—
	0,99	3.743	936	416	234	150	104	—	—	—	—
0,07	0,95	2.501	625	278	156	100	69	51	—	—	—
	0,99	4.320	1.080	480	270	173	120	88	—	—	—
0,08	0,95	2.827	707	314	177	113	79	58	44	—	—
	0,99	4.884	1.221	543	305	195	136	100	76	—	—
0,09	0,95	3.146	787	350	197	126	87	64	49	39	—
	0,99	5.435	1.359	604	340	217	151	111	85	67	—
0,10	0,95	3.457	864	384	216	138	96	71	54	43	35
	0,99	5.972	1.493	664	373	239	166	122	93	74	60
0,11	0,95	3.761	940	418	235	150	104	77	59	46	38
	0,99	6.496	1.624	722	406	260	180	133	102	80	65
0,12	0,95	4.057	1.014	451	254	162	113	83	63	50	41
	0,99	7.007	1.752	779	438	280	195	143	109	87	70
0,13	0,95	4.345	1.086	483	272	174	121	89	68	54	43
	0,99	7.505	1.876	834	469	300	208	153	117	93	75
0,14	0,95	4.625	1.156	514	289	185	128	94	72	57	46
	0,99	7.989	1.997	888	499	320	222	163	125	99	80
0,15	0,95	4.898	1.225	544	306	196	136	100	77	60	49
	0,99	8.461	2.115	940	529	338	235	173	132	104	85
0,16	0,95	5.163	1.291	574	323	207	143	105	81	64	52
	0,99	8.918	2.230	991	557	357	248	182	139	110	89
0,17	0,95	5.420	1.355	602	339	217	151	111	85	67	54
	0,99	9.363	2.341	1.135	585	375	260	191	146	116	94
0,18	0,95	5.670	1.418	630	354	227	158	116	89	70	57

(Continúa)

**Tabla A.** Número de sujetos necesarios para la estimación de una proporción (variable cualitativa) (cont.)

Proporción esperada (P)	Nivel de confianza (1- $\alpha$ )	Precisión (i)									
		0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,10
	0,99	9.794	2.449	1.088	612	392	272	200	153	121	98
0,19	0,95	5.912	1.478	657	370	236	164	121	92	73	59
	0,99	10.212	2.553	1.135	638	408	284	208	160	126	102
0,20	0,95	6.147	1.537	683	384	246	171	125	96	76	61
	0,99	10.617	2.654	1.180	664	425	295	217	166	131	106
0,25	0,95	7.203	1.801	800	450	288	200	147	113	89	72
	0,99	12.442	3.111	1.382	778	498	346	254	194	154	124
0,30	0,95	8.067	2.017	896	504	323	224	165	126	100	81
	0,99	13.935	3.484	1.548	871	557	387	284	218	172	139
0,35	0,95	8.740	2.185	971	546	350	243	178	137	108	87
	0,99	15.096	3.774	1.677	944	604	419	308	236	186	151
0,40	0,95	9.220	2.305	1.024	576	369	256	188	144	114	92
	0,99	15.926	3.981	1.770	995	637	442	325	249	197	159
0,45	0,95	9.508	2.377	1.056	594	380	264	194	149	117	95
	0,99	16.424	4.106	1.825	1.026	657	456	335	257	203	164
0,50	0,95	9.604	2.401	1.067	600	384	267	196	150	119	96
	0,99	16.589	4.147	1.843	1.037	664	461	339	259	205	166

- La fórmula utilizada para confeccionar esta tabla se fundamenta en la aproximación a la distribución normal y sólo es válida cuando  $n \cdot p$  y  $n \cdot (1-p)$  son iguales o superiores a 5.
- La tabla presenta los valores de P hasta 0,50, ya que es lo mismo estimar una proporción esperada del 30% de individuos con una determinada característica, que la del 70% de que no la tengan.

La fórmula para el cálculo del número de sujetos necesarios para estimar una proporción se encuentra en el cuadro 15.1. El ejemplo 15.1 ilustra su uso.

**Tabla B.** Número de sujetos necesarios para la estimación de una media (variable cuantitativa)

i/s*	Nivel de confianza (1 - $\alpha$ )		
	0,90	0,95	0,99
0,025	4.330	6.147	10.617
0,050	1.082	1.537	2.654
0,075	481	683	1.180
0,100	271	384	664
0,125	173	246	425
0,150	120	171	295
0,175	88	125	217

**Tabla B. Número de sujetos necesarios para la estimación de una media (variable cuantitativa) (cont.)**

i/s*	Nivel de confianza ( $1 - \alpha$ )		
	0,90	0,95	0,99
0,200	68	96	166
0,225	53	76	131
0,250	43	61	106
0,275	36	51	88
0,300	30	43	74
0,325	26	36	63
0,350	22	31	54
0,375	19	27	47
0,400	17	24	41
0,425	15	21	37
0,450	13	19	33
0,475	12	17	29
0,500	11	15	27
0,525	10	14	24
0,550	9	13	22
0,575	8	12	20
0,600	8	11	18
0,625	7	10	17
0,650	6	9	16
0,675	6	8	15
0,700	6	8	14
0,725	5	7	13
0,750	5	7	12
0,775	5	6	11
0,800	4	6	10
0,825	4	6	10
0,850	4	5	9
0,875	4	5	9
0,900	3	5	8
0,925	3	4	8
0,950	3	4	7
0,975	3	4	7
1,000	3	4	7

\*La fórmula para el cálculo del número de sujetos necesarios para estimar una media se encuentra en el cuadro 11.1. El ejemplo 11.2 ilustra su uso.

**Tabla C.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde a la proporción menor de las dos que se comparan

P1	P2	$\alpha$		Potencia ( $1 - \beta$ )			
		Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,05	0,10	0,050	0,10	342	474	598	871
		0,025	0,05	434	581	718	1.015
	0,15	0,050	0,10	110	152	192	279
		0,025	0,05	140	187	231	326
0,20	0,050	0,10	59	81	103	149	
		0,025	0,05	75	100	123	174
	0,25	0,050	0,10	38	53	66	96
		0,025	0,05	49	65	80	112
0,30	0,050	0,10	27	38	47	68	
		0,025	0,05	35	46	57	80
	0,35	0,050	0,10	21	29	36	51
		0,025	0,05	27	35	43	60
0,40	0,050	0,10	17	22	28	40	
		0,025	0,05	21	28	34	47
	0,45	0,050	0,10	13	18	23	32
		0,025	0,05	17	22	27	38
0,50	0,050	0,10	11	15	19	26	
		0,025	0,05	14	19	23	31
	0,55	0,050	0,10	9	12	15	22
		0,025	0,05	12	15	19	26
0,60	0,050	0,10	8	10	13	18	
		0,025	0,05	10	13	16	21
	0,65	0,050	0,10	7	9	11	15
		0,025	0,05	9	11	13	18
0,70	0,050	0,10	6	7	9	13	
		0,025	0,05	7	9	11	15
	0,75	0,050	0,10	5	6	8	10
		0,025	0,05	6	8	9	13
0,80	0,050	0,10	4	5	6	9	
		0,025	0,05	5	7	8	10
	0,85	0,050	0,10	4	4	5	7
		0,025	0,05	5	6	7	9
0,90	0,050	0,10	3	4	4	6	
		0,025	0,05	4	5	6	7

**Tabla C.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde a la proporción menor de las dos que se comparan (cont.)

P1	P2	$\alpha$		Potencia ( $1 - \beta$ )			
		Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,95	0,050	0,10	2	3	3	4	
	0,025	0,05	3	4	4	5	
	1,00	0,050	0,10	2	2	3	3
	0,025	0,05	3	3	3	4	
0,10	0,15	0,050	0,10	539	748	944	1.602
		0,025	0,05	685	918	1.134	397
	0,20	0,050	0,10	156	217	273	463
		0,025	0,05	199	266	328	198
0,25	0,050	0,10	78	108	136	231	
		0,025	0,05	99	133	164	1.375
	0,30	0,050	0,10	48	67	84	122
		0,025	0,05	62	82	101	142
0,35	0,050	0,10	33	46	58	83	
		0,025	0,05	43	57	70	97
	0,40	0,050	0,10	25	34	42	61
		0,025	0,05	31	42	51	71
0,45	0,050	0,10	19	26	32	47	
		0,025	0,05	24	32	39	55
	0,50	0,050	0,10	15	21	26	37
		0,025	0,05	19	25	31	43
0,55	0,050	0,10	12	17	21	29	
		0,025	0,05	16	21	25	35
	0,60	0,050	0,10	10	14	17	24
		0,025	0,05	13	17	21	28
0,65	0,050	0,10	8	11	14	20	
		0,025	0,05	11	14	17	23
	0,70	0,050	0,10	7	9	12	16
		0,025	0,05	9	12	14	19
0,75	0,050	0,10	6	8	10	13	
		0,025	0,05	8	10	12	16
	0,80	0,050	0,10	5	7	8	11
		0,025	0,05	7	8	10	13
0,85	0,050	0,10	4	6	7	9	
		0,025	0,05	6	7	8	11

**Tabla C.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde a la proporción menor de las dos que se comparan (cont.)

P1	P2	$\alpha$		Potencia ( $1 - \beta$ )			
		Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,90	0,050	0,10	4	5	5	7	
	0,025	0,05	5	6	7	9	
	0,95	0,050	0,10	3	4	4	6
	0,025	0,05	4	5	6	7	
0,95	1,00	0,050	0,10	2	3	3	4
	0,025	0,05	3	4	4	5	
	0,15	0,20	0,10	712	988	1.247	1.817
		0,025	0,05	904	1.212	1.498	2.117
0,95	0,25	0,050	0,10	197	272	344	500
		0,025	0,05	250	334	413	583
	0,30	0,050	0,10	95	131	165	240
	0,025	0,05	120	161	198	280	
0,95	0,35	0,050	0,10	57	78	99	143
		0,025	0,05	72	96	119	167
	0,40	0,050	0,10	38	53	66	96
	0,025	0,05	49	65	80	112	
0,95	0,45	0,050	0,10	28	38	48	69
		0,025	0,05	35	47	58	81
	0,50	0,050	0,10	21	29	36	52
	0,025	0,05	27	36	44	61	
0,95	0,55	0,050	0,10	16	22	28	40
		0,025	0,05	21	28	34	47
	0,60	0,050	0,10	13	18	22	32
	0,025	0,05	17	22	27	37	
0,95	0,65	0,050	0,10	11	14	18	26
		0,025	0,05	14	18	22	30
	0,70	0,050	0,10	9	12	15	21
	0,025	0,05	11	15	18	24	
0,95	0,75	0,050	0,10	7	10	12	17
		0,025	0,05	10	12	14	20
	0,80	0,050	0,10	6	8	10	14
	0,025	0,05	8	10	12	16	
0,95	0,85	0,050	0,10	5	7	8	11
		0,025	0,05	7	8	10	13

**Tabla C.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde a la proporción menor de las dos que se comparan (cont.)

P1	P2	$\alpha$		Potencia ( $1 - \beta$ )			
		Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,90	0,050	0,10	4	6	7	9	
	0,025	0,05	6	7	8	11	
	0,050	0,10	4	4	5	7	
	0,025	0,05	5	6	7	9	
0,95	0,050	0,10	4	4	5	7	
	0,025	0,05	5	6	7	9	
	0,050	0,10	3	4	4	5	
	0,025	0,05	4	5	5	7	
1,00	0,050	0,10	3	4	4	5	
	0,025	0,05	4	5	5	7	
	0,20	0,050	0,10	860	1.193	1.507	2.195
	0,025	0,05	1.093	1.464	1.810	2.558	
0,30	0,050	0,10	231	319	403	587	
	0,025	0,05	293	392	484	684	
	0,050	0,10	108	150	189	275	
	0,025	0,05	138	184	227	321	
0,35	0,050	0,10	64	88	111	161	
	0,025	0,05	81	108	133	188	
	0,40	0,050	0,10	64	88	111	161
	0,025	0,05	81	108	133	188	
0,45	0,050	0,10	42	58	73	106	
	0,025	0,05	54	72	88	124	
	0,50	0,050	0,10	30	41	52	75
	0,025	0,05	38	51	63	88	
0,55	0,050	0,10	23	31	39	56	
	0,025	0,05	39	38	47	65	
	0,60	0,050	0,10	17	24	30	43
	0,025	0,05	22	29	36	50	
0,65	0,050	0,10	14	19	23	33	
	0,025	0,05	18	23	28	39	
	0,70	0,050	0,10	11	15	19	26
	0,025	0,05	14	19	23	31	
0,75	0,050	0,10	9	12	15	21	
	0,025	0,05	12	15	18	25	
	0,80	0,050	0,10	7	10	12	17
	0,025	0,05	10	12	15	20	
0,85	0,050	0,10	6	8	10	14	
	0,025	0,05	8	10	12	16	
	0,90	0,050	0,10	5	7	8	11
	0,025	0,05	7	8	10	13	

**Tabla C.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde a la proporción menor de las dos que se comparan (*cont.*)

P1	P2	$\alpha$		Potencia ( $1 - \beta$ )			
		Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,95	0,050	0,10	4	5	6	9	
	0,025	0,05	5	7	8	10	
	1,00	0,050	0,10	3	4	5	7
	0,025	0,05	4	5	6	8	
0,25	0,30	0,050	0,10	984	1.365	1.724	2.511
		0,025	0,05	1.249	1.674	2.070	2.925
	0,35	0,050	0,10	258	358	452	658
		0,025	0,05	328	439	543	767
0,40	0,050	0,10	119	165	208	303	
		0,025	0,05	152	203	250	353
	0,45	0,050	0,10	69	96	120	175
		0,025	0,05	88	117	145	204
0,50	0,050	0,10	45	62	78	114	
		0,025	0,05	58	77	94	133
	0,55	0,050	0,10	32	44	55	79
		0,025	0,05	41	54	66	93
0,60	0,050	0,10	24	32	40	58	
		0,025	0,05	30	40	49	68
	0,65	0,050	0,10	18	25	31	44
		0,025	0,05	23	30	37	52
0,70	0,050	0,10	14	19	24	34	
		0,025	0,05	18	24	29	40
	0,75	0,050	0,10	11	15	19	27
		0,025	0,05	14	19	23	32
0,80	0,050	0,10	9	12	15	21	
		0,025	0,05	12	15	18	25
	0,85	0,050	0,10	7	10	12	17
		0,025	0,05	10	12	15	20
0,90	0,050	0,10	6	8	10	13	
		0,025	0,05	8	10	12	16
	0,95	0,050	0,10	5	6	8	10
		0,025	0,05	6	8	9	13
1,00	0,050	0,10	4	5	6	8	
		0,025	0,05	5	6	8	10

**Tabla C.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde a la proporción menor de las dos que se comparan (cont.)

P1	P2	α		Potencia (1 – β)			
		Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,30	0,35	0,025	0,10	1.083	1.502	1.897	2.763
		0,050	0,05	1.375	1.843	2.278	3.219
	0,40	0,025	0,10	280	388	490	713
		0,050	0,05	356	476	588	831
	0,45	0,025	0,10	128	177	223	324
		0,050	0,05	162	217	268	278
	0,50	0,025	0,10	73	101	127	185
		0,050	0,05	93	124	153	215
0,55	0,55	0,025	0,10	47	65	82	119
		0,050	0,05	60	80	99	139
	0,60	0,025	0,10	33	45	57	82
		0,050	0,05	42	56	68	96
	0,65	0,025	0,10	24	33	41	60
		0,050	0,05	31	41	50	70
	0,70	0,025	0,10	18	25	31	45
		0,050	0,05	23	31	38	52
0,75	0,75	0,025	0,10	14	19	24	34
		0,050	0,05	18	24	29	40
	0,80	0,025	0,10	11	15	19	26
		0,050	0,05	14	19	23	31
	0,85	0,025	0,10	9	12	15	21
		0,050	0,05	11	15	18	24
	0,90	0,025	0,10	7	9	12	16
		0,050	0,05	9	12	14	19
0,95	0,95	0,025	0,10	6	7	9	13
		0,050	0,05	7	9	11	15
	1,00	0,025	0,10	5	6	7	10
		0,025	0,05	6	7	9	12
	0,35	0,40	0,10	1.157	1.604	2.027	2.952
		0,25	0,05	1.469	1.969	2.434	3.439
	0,45	0,50	0,10	295	409	517	752
		0,25	0,05	375	502	621	877
	0,50	0,50	0,10	133	184	232	338
		0,25	0,05	169	226	279	394

**Tabla C.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde a la proporción menor de las dos que se comparan (*cont.*)

P1	P2	$\alpha$		Potencia ( $1 - \beta$ )			
		Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,55	0,050	0,10	75	104	131	191	
	0,025	0,05	96	128	158	222	
0,60	0,050	0,10	48	66	84	121	
	0,025	0,05	61	81	101	142	
0,65	0,050	0,10	33	46	57	83	
	0,025	0,05	42	56	69	97	
0,70	0,050	0,10	24	33	41	60	
	0,025	0,05	31	41	50	70	
0,75	0,025	0,05	18	25	31	44	
	0,050	0,10	23	30	37	52	
0,80	0,025	0,05	14	19	23	33	
	0,050	0,10	18	23	28	39	
0,85	0,025	0,05	11	14	18	26	
	0,050	0,10	14	18	22	30	
0,90	0,025	0,05	8	11	14	20	
	0,050	0,10	11	14	17	23	
0,95	0,025	0,05	7	9	11	15	
	0,050	0,10	9	11	13	18	
1,00	0,025	0,05	5	7	8	11	
	0,025	0,05	7	9	10	14	
0,40	0,45	0,050	0,10	1.206	1.673	2.113	3.078
		0,025	0,05	1.532	2.053	2.538	3.586
0,50	0,50	0,050	0,10	305	422	533	776
		0,025	0,05	387	518	640	904
0,55	0,55	0,050	0,10	136	188	237	345
		0,025	0,05	173	231	285	402
0,60	0,60	0,050	0,10	76	105	133	192
		0,025	0,05	97	129	159	225
0,65	0,65	0,050	0,10	48	66	84	121
		0,025	0,05	61	82	101	142
0,70	0,70	0,050	0,10	33	45	57	82
		0,025	0,05	42	56	68	96
0,75	0,75	0,050	0,10	24	32	40	58
		0,025	0,05	30	40	49	68

**Tabla C.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde a la proporción menor de las dos que se comparan (cont.)

P1	P2	$\alpha$		Potencia ( $1 - \beta$ )			
		Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,80	0,050	0,10	17	24	30	43	
	0,025	0,05	22	29	36	50	
0,85	0,050	0,10	13	18	22	32	
	0,025	0,05	17	22	27	37	
0,90	0,050	0,10	10	14	17	24	
	0,025	0,05	13	17	21	28	
0,95	0,050	0,10	8	10	13	18	
	0,025	0,05	10	13	16	21	
1,00	0,050	0,10	6	8	10	14	
	0,025	0,05	8	10	12	16	
0,45	0,50	0,10	1.231	1.707	2.157	3.141	
	0,025	0,05	1.563	2.095	2.590	3.660	
0,55	0,050	0,10	308	426	538	784	
	0,025	0,05	391	523	647	913	
0,60	0,050	0,10	136	188	237	345	
	0,025	0,05	173	231	285	402	
0,65	0,050	0,10	75	104	131	191	
	0,025	0,05	96	128	158	222	
0,70	0,050	0,10	47	65	82	119	
	0,025	0,05	60	80	99	139	
0,75	0,050	0,10	32	44	55	79	
	0,025	0,05	41	54	66	93	
0,80	0,050	0,10	23	31	39	56	
	0,025	0,05	29	38	47	65	
0,85	0,050	0,10	16	22	28	40	
	0,025	0,05	21	28	34	47	
0,90	0,050	0,10	12	17	21	29	
	0,025	0,05	16	21	25	35	
0,95	0,050	0,10	9	12	15	22	
	0,025	0,05	12	15	19	26	
1,00	0,050	0,10	7	9	11	16	
	0,025	0,05	9	12	14	19	
0,50	0,55	0,10	1.231	1.707	2.157	3.141	
	0,025	0,05	1.563	2.095	2.590	3.660	

**Tabla C.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde a la proporción menor de las dos que se comparan (cont.)

P1	P2	$\alpha$		Potencia ( $1 - \beta$ )			
		Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,60	0,050	0,10	305	422	533	776	
	0,025	0,05	387	518	640	904	
0,65	0,050	0,10	133	184	232	338	
	0,025	0,05	169	226	279	394	
0,70	0,050	0,10	73	101	127	185	
	0,025	0,05	93	124	153	215	
0,75	0,050	0,10	44	59	74	106	
	0,025	0,05	56	73	90	125	
0,80	0,050	0,10	30	41	52	75	
	0,025	0,05	38	51	63	88	
0,85	0,050	0,10	21	29	36	52	
	0,025	0,05	27	36	44	61	
0,90	0,050	0,10	15	21	26	37	
	0,025	0,05	19	25	31	43	
0,95	0,050	0,10	11	15	19	26	
	0,025	0,05	14	19	23	31	
1,00	0,050	0,10	8	11	13	19	
	0,025	0,05	11	14	16	22	
0,55	0,60	0,050	0,05	1.206	1.673	2.113	3.078
		0,025	0,10	1.532	2.053	2.538	3.586
0,65	0,050	0,05	295	409	517	752	
	0,025	0,10	375	502	621	877	
0,70	0,050	0,05	128	177	223	324	
	0,025	0,10	162	217	268	378	
0,75	0,050	0,05	69	96	120	175	
	0,025	0,10	88	117	145	204	
0,80	0,050	0,05	42	58	73	106	
	0,025	0,10	54	72	88	124	
0,85	0,050	0,05	28	38	48	69	
	0,025	0,10	35	47	58	81	
0,90	0,050	0,05	19	26	32	47	
	0,025	0,10	24	32	39	55	
0,95	0,050	0,05	13	18	23	32	
	0,025	0,10	17	22	27	38	

**Tabla C.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde a la proporción menor de las dos que se comparan (cont.)

P1	P2	$\alpha$		Potencia ( $1 - \beta$ )			
		Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,60	1,00	0,050	0,05	10	13	16	22
		0,025	0,05	12	16	19	26
	0,65	0,050	0,10	1.157	1.604	2.027	2.952
		0,025	0,05	1.469	1.969	2.434	3.439
0,70	0,70	0,050	0,10	280	388	490	713
		0,025	0,05	356	476	588	831
	0,75	0,050	0,10	119	165	208	303
		0,025	0,05	152	203	250	353
0,80	0,80	0,050	0,10	64	88	111	161
		0,025	0,05	81	108	133	188
	0,85	0,050	0,10	38	53	66	96
		0,025	0,05	49	65	80	112
0,90	0,90	0,050	0,10	25	34	42	61
		0,025	0,05	31	42	51	71
	0,95	0,050	0,10	17	22	28	40
		0,025	0,05	21	28	34	47
0,65	1,00	0,050	0,10	11	15	19	27
		0,025	0,05	14	19	23	32
	0,70	0,050	0,10	1.083	1.502	1.897	2.763
		0,025	0,05	1.375	1.843	2.278	3.219
0,75	0,75	0,050	0,10	258	358	452	658
		0,025	0,05	328	439	543	767
	0,80	0,050	0,10	108	150	189	275
		0,025	0,05	138	184	227	321
0,85	0,85	0,050	0,10	57	78	99	143
		0,025	0,05	72	96	119	167
	0,90	0,050	0,10	33	46	58	83
		0,025	0,05	43	57	70	97
0,95	0,95	0,050	0,10	21	29	36	51
		0,025	0,05	27	35	43	60
	1,00	0,050	0,10	13	18	23	32
		0,025	0,05	17	23	28	38
0,70	0,75	0,050	0,10	984	1.365	1.724	2.511
		0,025	0,05	1.249	1.674	2.070	2.925

**Tabla C.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde a la proporción menor de las dos que se comparan (cont.)

P1	P2	$\alpha$		Potencia ( $1 - \beta$ )			
		Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,80	0,050	0,10	231	319	403	587	
	0,025	0,05	293	392	484	684	
	0,050	0,10	95	131	165	240	
	0,025	0,05	120	161	198	280	
0,85	0,050	0,10	48	67	84	122	
	0,025	0,05	62	82	101	142	
	0,050	0,10	27	38	47	68	
	0,025	0,05	35	46	57	80	
0,90	0,050	0,10	16	22	28	40	
	0,025	0,05	21	28	34	47	
	0,050	0,10	860	1.193	1.507	2.195	
	0,025	0,05	1.093	1.464	1.810	2.558	
0,95	0,050	0,10	197	272	344	500	
	0,025	0,05	250	334	413	583	
	0,050	0,10	78	108	136	198	
	0,025	0,05	99	133	164	231	
1,00	0,050	0,10	38	53	66	96	
	0,025	0,05	49	65	80	112	
	0,050	0,10	21	28	35	50	
	0,025	0,05	26	35	42	59	
0,75	0,80	0,10	712	988	1.247	1.817	
	0,025	0,05	904	1.212	1.498	2.117	
	0,050	0,10	156	217	273	397	
	0,025	0,05	199	266	328	463	
0,80	0,85	0,10	59	81	103	149	
	0,025	0,05	75	100	123	174	
	0,050	0,10	27	37	46	66	
	0,025	0,05	34	45	55	78	
0,85	0,90	0,10	539	748	944	1.375	
	0,025	0,05	685	918	1.134	1.602	
	0,050	0,10	110	152	192	279	
	0,025	0,05	140	187	231	326	
1,00	0,050	0,10	37	51	64	93	
	0,025	0,05	47	63	77	108	

**Tabla C.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos proporciones. P1 corresponde a la proporción menor de las dos que se comparan (cont.)

P1	P2	$\alpha$		Potencia (1 – $\beta$ )			
		Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,90	0,91	0,050	0,10	10.617	14.730	18.609	27.110
		0,025	0,05	13.480	18.071	22.344	31.582
	0,92	0,050	0,10	2.528	3.506	4.430	6.453
		0,025	0,05	3.209	4.302	5.319	7.517
0,93	0,93	0,050	0,10	1.066	1.479	1.868	2.721
		0,025	0,05	1.354	1.814	2.243	3.170
	0,94	0,050	0,10	567	786	993	1.446
		0,025	0,05	720	965	1.193	1.685
0,95	0,95	0,050	0,10	342	474	598	871
		0,025	0,05	434	581	718	1.015
	1,00	0,050	0,10	58	79	100	145
		0,025	0,05	73	98	120	169
0,95	0,96	0,050	0,10	5.307	7.362	9.301	9.301
		0,025	0,05	6.737	9.032	11.167	11.167
	0,97	0,050	0,10	1.185	1.643	2.076	2.076
		0,025	0,05	1.504	2.016	2.492	2.492
0,98	0,98	0,050	0,10	462	641	810	810
		0,025	0,05	587	787	972	972
	0,99	0,050	0,10	224	310	391	391
		0,025	0,05	284	380	470	470
1,00	1,00	0,050	0,10	119	165	208	208
		0,025	0,05	152	203	250	250

La fórmula para el cálculo del número de sujetos necesarios para comparar dos porciones se encuentra en el cuadro 15.3. El ejemplo 15.4 ilustra su uso.

**Tabla D.** Número de sujetos necesarios en cada grupo de estudio para estimar un riesgo relativo (RR)

P1	$\alpha$		Potencia 0,80					
	Hipótesis unilateral	Hipótesis bilateral	1,5	2,0	RR 2,5	3,0	3,5	4,0
0,01	0,050	0,10	6.107	1.827	944	605	434	334
	0,025	0,05	7.752	2.319	1.199	768	551	424
0,02	0,050	0,10	3.014	899	463	296	212	162
	0,025	0,05	3.826	1.141	588	376	269	206

(Continúa)

**Tabla D.** Número de sujetos necesarios en cada grupo de estudio para estimar un riesgo relativo (RR) (cont.)

P1	$\alpha$		Potencia 0,80					
	Hipótesis unilateral	Hipótesis bilateral	1,5	2,0	RR 2,5	3,0	3,5	4,0
0,03	0,050	0,10	1.983	590	303	193	137	105
	0,025	0,05	2.518	749	385	245	175	133
0,04	0,050	0,10	1.468	435	223	141	100	76
	0,025	0,05	1.863	552	283	179	127	97
0,05	0,050	0,10	1.159	342	175	110	78	59
	0,025	0,05	1.471	435	222	140	99	75
0,06	0,050	0,10	952	280	142	90	63	48
	0,025	0,05	1.209	356	181	114	80	61
0,07	0,050	0,10	805	236	120	75	53	39
	0,025	0,05	1.022	300	152	95	67	50
0,08	0,050	0,10	695	203	102	64	45	33
	0,025	0,05	882	258	130	81	57	42
0,09	0,050	0,10	609	177	89	55	38	29
	0,025	0,05	773	225	113	70	49	36
0,10	0,050	0,10	540	157	78	48	33	25
	0,025	0,05	686	199	100	62	43	32
0,15	0,050	0,10	334	95	46	28	19	13
	0,025	0,05	424	121	59	35	24	17
0,20	0,050	0,10	231	64	30	17	11	7
	0,025	0,05	293	81	38	22	14	10
0,25	0,050	0,10	169	45	21	11	7	4
	0,025	0,05	215	58	26	14	9	5
0,30	0,050	0,10	128	33	14	7	4	–
	0,025	0,05	162	42	18	9	5	–
0,35	0,050	0,10	98	24	10	4	–	–
	0,025	0,05	125	31	12	5	–	–
0,40	0,050	0,10	76	17	6	–	–	–
	0,025	0,05	97	22	8	–	–	–
0,45	0,050	0,10	59	12	3	–	–	–
	0,025	0,05	75	16	4	–	–	–
0,50	0,050	0,10	45	8	–	–	–	–
	0,025	0,05	58	11	–	–	–	–

**Tabla D.** Número de sujetos necesarios en cada grupo de estudio para estimar un riesgo relativo (RR) (cont.)

P1	$\alpha$		Potencia 0,90					
	Hipótesis unilateral	Hipótesis bilateral	1,5	2,0	RR 2,5	3,0	3,5	4,0
0,01	0,050	0,10	8.458	2.530	1.307	838	601	462
	0,025	0,05	10.377	3.104	1.604	1.028	738	567
0,02	0,050	0,10	4.175	1.245	641	409	293	224
	0,025	0,05	5.122	1.527	787	502	359	275
0,03	0,050	0,10	2.747	816	419	267	190	145
	0,025	0,05	3.370	1.002	514	327	233	178
0,04	0,050	0,10	2.033	602	308	195	138	105
	0,025	0,05	2.494	739	378	240	170	129
0,05	0,050	0,10	1.604	474	241	152	108	129
	0,025	0,05	1.969	581	296	187	132	81
0,06	0,050	0,10	1.319	388	197	124	87	100
	0,025	0,05	1.618	476	242	152	107	66
0,07	0,050	0,10	1.115	327	165	103	72	81
	0,025	0,05	1.368	401	203	127	89	54
0,08	0,050	0,10	962	281	141	88	61	67
	0,025	0,05	1.180	345	174	108	75	46
0,09	0,050	0,10	843	245	123	76	53	56
	0,025	0,05	1.034	301	151	94	65	39
0,10	0,050	0,10	748	217	108	67	46	48
	0,025	0,05	918	266	133	82	57	34
0,15	0,050	0,10	462	131	64	38	25	42
	0,025	0,05	567	161	78	47	31	18
0,20	0,050	0,10	319	88	41	24	15	22
	0,025	0,05	392	108	51	29	19	10
0,25	0,050	0,10	234	62	28	15	9	12
	0,025	0,05	287	77	35	19	11	5
0,30	0,050	0,10	177	45	19	9	5	6
	0,025	0,05	217	56	24	12	6	—
0,35	0,050	0,10	136	33	13	5	—	—
	0,025	0,05	167	41	16	7	—	—
0,40	0,050	0,10	105	24	8	—	—	—
	0,025	0,05	129	29	10	—	—	—
0,45	0,050	0,10	81	17	4	—	—	—
	0,025	0,05	100	21	5	—	—	—

**Tabla D.** Número de sujetos necesarios en cada grupo de estudio para estimar un riesgo relativo (RR) (cont.)

P1	α		Potencia 0,90					
	Hipótesis unilateral	Hipótesis bilateral	1,5	2,0	RR 2,5	3,0	3,5	4,0
0,50	0,050	0,10	62	11	—	—	—	—
	0,025	0,05	77	14	—	—	—	—

P1: riesgo de desarrollar el desenlace en el grupo no expuesto.

RR: magnitud mínima del riesgo relativo (RR) que se desea ser capaz de detectar.

La fórmula para el cálculo del número de sujetos necesarios para estimar un RR es la misma que la de comparación de dos proporciones (cuadro 15.3) en la que P2 = P1·RR. El ejemplo 15.5 ilustra su uso.

**Tabla E.** Número de sujetos necesarios en cada grupo de estudio para estimar una odds ratio (OR)

P1	α		Potencia 0,80					
	Hipótesis unilateral	Hipótesis bilateral	1,5	2,0	OR 2,5	3,0	3,5	4,0
0,01	0,050	0,10	6.275	1.889	982	633	457	353
	0,025	0,05	7.965	2.398	1.247	804	580	449
0,02	0,050	0,10	3.183	962	501	324	234	182
	0,025	0,05	4.041	1.221	637	411	298	231
0,03	0,050	0,10	2.153	653	341	221	160	125
	0,025	0,05	2.733	829	433	281	204	158
0,04	0,050	0,10	1.639	498	261	170	123	96
	0,025	0,05	2.080	633	332	216	157	122
0,05	0,050	0,10	1.331	406	214	139	101	79
	0,025	0,05	1.689	516	271	177	129	100
0,06	0,050	0,10	1.125	345	182	119	87	68
	0,025	0,05	1.429	438	231	151	110	86
0,07	0,050	0,10	979	301	159	104	76	60
	0,025	0,05	1.243	382	202	132	97	76
0,08	0,050	0,10	870	268	142	93	68	54
	0,025	0,05	1.104	341	181	119	87	68
0,09	0,050	0,10	785	243	129	85	62	49
	0,025	0,05	997	308	164	108	79	63
0,10	0,050	0,10	717	223	119	78	58	45
	0,025	0,05	911	283	151	100	73	58
0,15	0,050	0,10	517	163	88	59	44	35
	0,025	0,05	657	207	112	75	56	44
0,20	0,050	0,10	421	135	74	50	38	30
	0,025	0,05	535	172	94	64	48	38

**Tabla E.** Número de sujetos necesarios en cada grupo de estudio para estimar una odds ratio (OR) (cont.)

P1	$\alpha$		Potencia 0,80					
	Hipótesis unilateral	Hipótesis bilateral	1,5	2,0	OR 2,5	3,0	3,5	4,0
0,25	0,050	0,10	367	120	66	45	34	28
	0,025	0,05	466	152	84	58	44	36
0,30	0,050	0,10	335	111	62	43	33	27
	0,025	0,05	425	141	79	55	42	34
0,35	0,050	0,10	315	106	60	42	33	27
	0,025	0,05	400	135	77	54	41	34
0,40	0,050	0,10	305	104	60	42	33	27
	0,025	0,05	387	132	76	54	42	35
0,45	0,050	0,10	302	105	61	43	34	28
	0,025	0,05	383	133	78	55	43	36
0,50	0,050	0,10	305	107	63	45	36	30
	0,025	0,05	387	136	80	58	46	38
$\alpha$		Potencia 0,90						
P1	Hipótesis unilateral	Hipótesis bilateral	1,5	2,0	OR 2,5	3,0	3,5	4,0
0,01	0,025	0,10	8.691	2.616	1.360	877	632	489
	0,050	0,05	10.662	3.210	1.669	1.076	776	600
0,02	0,025	0,10	4.409	1.331	694	448	324	251
	0,050	0,05	5.409	1.634	852	550	398	308
0,03	0,025	0,10	2.982	904	472	306	222	172
	0,050	0,05	3.659	1.109	580	376	272	211
0,04	0,025	0,10	2.270	690	362	235	171	133
	0,025	0,05	2.785	847	444	288	209	163
0,05	0,050	0,10	1.843	562	295	192	140	109
	0,025	0,05	2.261	690	363	236	172	134
0,06	0,050	0,10	1.558	477	251	164	120	93
	0,025	0,05	1.912	585	309	201	147	115
0,07	0,050	0,10	1.356	416	220	144	105	82
	0,025	0,05	1.664	511	270	177	129	101
0,08	0,050	0,10	1.204	371	197	129	94	74
	0,025	0,05	1.478	455	241	158	116	91
0,09	0,050	0,10	1.087	336	179	117	86	68
	0,025	0,05	1.334	412	219	144	106	83

## Métodos de investigación clínica y epidemiológica

**Tabla E.** Número de sujetos necesarios en cada grupo de estudio para estimar una *odds ratio* (OR) (cont.)

P1	$\alpha$		Potencia 0,90					
	Hipótesis unilateral	Hipótesis bilateral	1,5	2,0	OR 2,5	3,0	3,5	4,0
0,10	0,050	0,10	993	308	164	108	79	63
	0,025	0,05	1.219	378	202	133	98	77
0,15	0,050	0,10	716	226	122	81	60	48
	0,025	0,05	879	277	150	100	74	59
0,20	0,050	0,10	583	187	102	69	52	41
	0,025	0,05	715	229	125	85	64	51
0,25	0,050	0,10	508	165	92	62	47	38
	0,025	0,05	623	203	113	77	58	47
0,30	0,050	0,10	463	153	86	59	45	37
	0,025	0,05	568	188	106	73	56	45
0,35	0,050	0,10	436	146	83	58	45	37
	0,025	0,05	536	180	102	71	55	45
0,40	0,050	0,10	422	144	83	58	45	37
	0,025	0,05	518	177	102	71	56	46
0,45	0,050	0,10	418	144	84	60	47	39
	0,025	0,05	513	177	103	73	58	48
0,50	0,050	0,10	422	148	87	62	49	41
	0,025	0,05	518	182	107	77	61	51

P1: proporción de exposición esperada entre los controles.

OR:magnitud mínima de la *odds ratio* (OR) que se desea ser capaz de detectar.

La fórmula para el cálculo del número de sujetos necesarios para estimar una OR es la misma que la de comparación de dos proporciones (cuadro 15.3) en la que:

$$P2 = (P1 \cdot OR) / [1 + P1 \cdot (OR - 1)].$$

El ejemplo 15.6 ilustra su uso.

**Tabla F.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos medias mediante la prueba de la t de Student-Fisher

d/s*	$\alpha$		Potencia (1 – $\beta$ )			
	Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,025	0,050	0,10	19.761	27.415	34.637	50.460
	0,025	0,05	25.088	33.634	41.587	58.783
0,050	0,050	0,10	4.940	6.854	8.659	12.615
	0,025	0,05	6.272	8.408	10.397	14.696
0,075	0,050	0,10	2.196	3.046	3.849	5.607
	0,025	0,05	2.788	3.737	4.621	6.531

**Tabla F.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos medias mediante la prueba de la t de Student-Fisher (cont.).

d/s*	$\alpha$		Potencia ( $1 - \beta$ )			
	Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,100	0,050	0,10	1.235	1.713	2.165	3.154
	0,025	0,05	1.568	2.102	2.599	3.674
0,125	0,050	0,10	790	1.097	1.385	2.018
	0,025	0,05	1.004	1.345	1.663	2.351
0,150	0,050	0,10	549	762	962	1.402
	0,025	0,05	697	934	1.155	1.633
0,175	0,050	0,10	403	559	707	1.030
	0,025	0,05	512	686	849	1.200
0,200	0,050	0,10	309	428	541	788
	0,025	0,05	392	526	650	918
0,225	0,050	0,10	244	338	428	623
	0,025	0,05	310	415	513	726
0,250	0,050	0,10	198	274	346	505
	0,025	0,05	251	336	416	588
0,275	0,050	0,10	163	227	286	417
	0,025	0,05	207	278	344	486
0,300	0,050	0,10	137	190	241	250
	0,025	0,05	174	234	289	408
0,325	0,050	0,10	117	162	205	299
	0,025	0,05	148	199	246	248
0,350	0,050	0,10	101	140	177	257
	0,025	0,05	128	172	212	300
0,375	0,050	0,10	88	122	154	224
	0,025	0,05	112	149	185	261
0,400	0,050	0,10	77	107	135	197
	0,025	0,05	98	131	162	230
0,425	0,050	0,10	68	95	120	175
	0,025	0,05	87	116	144	203
0,450	0,050	0,10	61	85	107	156
	0,025	0,05	77	104	128	181
0,475	0,050	0,10	55	76	96	140
	0,025	0,05	69	93	115	163
0,500	0,050	0,10	49	69	87	126
	0,025	0,05	63	84	104	147

**Tabla F.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos medias mediante la prueba de la t de Student-Fisher (cont.)

d/s*	$\alpha$		Potencia ( $1 - \beta$ )			
	Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,525	0,050	0,10	45	62	79	114
	0,025	0,05	57	76	94	133
0,550	0,050	0,10	41	57	72	104
	0,025	0,05	52	69	86	121
0,575	0,050	0,10	37	52	65	95
	0,025	0,05	47	64	79	111
0,600	0,050	0,10	34	48	60	88
	0,025	0,05	44	58	72	102
0,625	0,050	0,10	32	44	55	81
	0,025	0,05	40	54	67	94
0,650	0,050	0,10	29	41	51	75
	0,025	0,05	37	50	62	87
0,675	0,050	0,100	27	38	48	69
	0,025	0,050	34	46	57	81
0,700	0,050	0,100	25	35	44	64
	0,025	0,050	32	43	53	75
0,725	0,050	0,100	23	33	41	60
	0,025	0,050	30	40	49	70
0,750	0,050	0,100	22	30	38	56
	0,025	0,050	28	37	46	65
0,775	0,050	0,100	21	29	36	53
	0,025	0,050	26	35	43	61
0,800	0,050	0,100	19	27	34	49
	0,025	0,050	25	33	41	57
0,825	0,050	0,100	18	25	32	46
	0,025	0,050	23	31	38	54
0,850	0,050	0,100	17	24	30	44
	0,025	0,050	22	29	36	51
0,875	0,050	0,100	16	22	28	41
	0,025	0,050	20	27	34	48
0,900	0,050	0,100	15	21	27	39
	0,025	0,050	19	26	32	45
0,925	0,050	0,100	14	20	25	37
	0,025	0,050	18	25	30	43

**Tabla F.** Número de sujetos necesarios en cada grupo de estudio para la comparación de dos medias mediante la prueba de la t de Student-Fisher (cont.).

d/s*	$\alpha$		Potencia (1 – $\beta$ )			
	Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,950	0,050	0,100	14	19	24	35
	0,025	0,050	17	23	29	41
0,975	0,050	0,100	13	18	23	33
	0,025	0,050	16	22	27	39
1,000	0,050	0,100	12	17	22	32
	0,025	0,050	16	21	26	37
1,025	0,050	0,100	12	16	21	30
	0,025	0,050	15	20	25	35
1,050	0,050	0,100	11	16	20	29
	0,025	0,050	14	19	24	33
1,075	0,050	0,100	11	15	19	27
	0,025	0,050	14	18	22	32
1,100	0,050	0,100	10	14	18	26
	0,025	0,050	13	17	21	30

\*d/s es una medida estandarizada de la diferencia mínima que se desea detectar. Se calcula dividiendo la magnitud de la diferencia (d) por la desviación estándar de la variable en el grupo de referencia (s).

La fórmula para el cálculo del número de sujetos necesarios para comparar dos medias se encuentra en el cuadro 15.3. El ejemplo 15.7 ilustra su uso.

**Tabla G.** Número de sujetos necesarios en cada grupo en estudios de equivalencia

P	$\alpha$		Potencia = 0,80			Potencia = 0,90		
	Hipótesis unilateral	Hipótesis bilateral	0,05	d – d*	0,10	0,05	d – d*	0,10
0,05	0,050	0,10	235	59	26	326	81	36
	0,025	0,05	298	75	33	399	100	44
0,10	0,050	0,10	445	111	49	617	154	69
	0,025	0,05	565	141	63	757	189	84
0,15	0,050	0,10	631	158	70	874	218	97
	0,025	0,05	801	200	89	1.072	268	119
0,20	0,050	0,10	792	198	88	1.097	274	122
	0,025	0,05	1.005	251	112	1.345	336	149
0,25	0,050	0,10	928	232	103	1.285	321	143
	0,025	0,05	1.178	294	131	1.577	394	175
0,30	0,050	0,10	1.039	260	115	1.439	360	160
	0,025	0,05	1.319	330	147	1.766	441	196

**Tabla G.** Número de sujetos necesarios en cada grupo en estudios de equivalencia (cont.)

P	$\alpha$	Potencia = 0,80			Potencia = 0,90			
		Hipótesis unilateral	Hipótesis bilateral	0,05	d – d*	0,10	0,15	d – d*
0,35	0,050	0,10	1.126	281	125	1.559	390	173
	0,025	0,05	1.429	357	159	1.913	478	213
0,40	0,050	0,10	1.188	297	132	1.645	411	183
	0,025	0,05	1.507	377	167	2.018	505	224
0,45	0,050	0,10	1.225	306	136	1.696	424	188
	0,025	0,05	1.555	389	173	2.081	520	231
0,50	0,050	0,10	1.237	309	137	1.713	428	190
	0,025	0,05	1.570	393	174	2.102	526	234
0,55	0,050	0,10	1.225	306	136	1.696	424	188
	0,025	0,05	1.555	389	173	2.081	520	231
0,60	0,050	0,10	1.188	297	132	1.645	411	183
	0,025	0,05	1.507	377	167	2.018	505	224
0,65	0,050	0,10	1.126	281	125	1.559	390	173
	0,025	0,05	1.429	357	159	1.913	478	213
0,70	0,050	0,10	1.039	260	115	1.439	360	160
	0,025	0,05	1.319	330	147	1.766	441	196
0,75	0,050	0,10	928	232	103	1.285	321	143
	0,025	0,05	1.178	294	131	1.577	394	175
0,80	0,050	0,10	792	198	88	1.097	274	122
	0,025	0,05	1.005	251	112	1.345	336	149
0,85	0,050	0,10	631	158	70	874	218	97
	0,025	0,05	801	200	89	1.072	268	119
0,90	0,050	0,10	445	111	49	617	154	69
	0,025	0,05	565	141	63	757	189	84
0,95	0,050	0,10	235	59	26	326	81	36
	0,025	0,05	298	75	33	399	100	44

P: porcentaje esperado en el grupo de referencia.

d: diferencia máxima entre las respuestas a ambos tratamientos que se tolerará para concluir que son equivalentes (margen de equivalencia).

La fórmula para el cálculo del número de sujetos necesarios en un estudio de equivalencia se encuentra en el cuadro 15.4.

El ejemplo 15.8 ilustra su uso.

\*d: diferencia esperada entre las respuestas a ambos tratamientos.

**Tabla H.** Número de sujetos necesarios para la estimación de la concordancia intraobservador

Amplitud del intervalo de confianza del 95% (2i)						
e	0,025	0,050	0,075	0,100	0,125	0,150
0,025	316	79	—	—	—	—
0,050	653	164	73	41	—	—
0,075	1.017	255	113	64	41	29
0,100	1.418	355	158	89	57	40
0,125	1.868	467	208	117	75	52
0,150	2.384	596	265	149	95	67

e: porcentaje de error o discordancia entre ambas mediciones que se espera encontrar. El ejemplo A7.3 ilustra su uso.

**Tabla I.** Número de sujetos necesarios para la estimación de la concordancia intra e interobservador

Amplitud del intervalo de confianza del 95% (2i)						
p (dis)	e	0,050	0,100	0,150	0,200	0,250
0,05	0,05	416	104	—	—	—
	0,10	775	194	—	—	—
	0,15	1.686	422	—	—	—
0,10	0,05	731	183	81	46	—
	0,10	1.142	285	127	71	—
	0,15	2.078	519	231	130	—
0,15	0,05	1.009	252	112	63	40
	0,10	1.465	366	163	92	59
	0,15	2.423	606	269	151	97
0,20	0,05	1.250	313	139	78	50
	0,10	1.746	436	194	79	70
	0,15	2.723	681	303	170	109
0,25	0,05	1.455	364	162	91	58
	0,10	1.983	496	220	124	79
	0,15	2.976	744	331	186	119
0,30	0,05	1.621	405	180	101	65
	0,10	2.177	544	242	136	87
	0,15	3.184	796	354	199	127

p (dis): porcentaje de discordancia esperado entre ambos observadores.

e: porcentaje de discordancia esperado dentro de cada observador (se asume que la probabilidad de error intraobservador es la misma para ambos observadores).

El ejemplo A7.4 ilustra su uso.

**Tabla J.** Número de sujetos necesarios para estimar una asociación entre dos variables cuantitativas mediante el coeficiente de correlación de Pearson

$r^*$	$\alpha$		Potencia ( $1 - \beta$ )			
	Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,05	0,050	0,10	2.473	3.424	4.325	6.300
	0,025	0,05	3.138	4.200	5.193	7.339
0,10	0,050	0,10	617	854	1.078	1.569
	0,025	0,05	783	1.047	1.294	1.828
0,15	0,050	0,10	274	378	477	693
	0,025	0,05	347	463	572	807
0,20	0,050	0,10	153	211	266	387
	0,025	0,05	194	259	319	450
0,25	0,050	0,10	98	134	169	245
	0,025	0,05	123	164	202	285
0,30	0,050	0,10	68	92	116	168
	0,025	0,05	85	113	139	195
0,35	0,050	0,10	49	67	84	121
	0,025	0,05	62	82	100	141
0,40	0,050	0,10	37	51	63	91
	0,025	0,05	47	62	75	105
0,45	0,050	0,10	29	39	49	70
	0,025	0,05	36	48	58	81
0,50	0,050	0,10	23	31	39	55
	0,025	0,05	29	38	46	64
0,55	0,050	0,10	19	25	31	44
	0,025	0,05	24	30	37	51
0,60	0,050	0,10	16	21	26	36
	0,025	0,05	19	25	30	41
0,65	0,050	0,10	13	17	21	29
	0,025	0,05	16	20	25	34
0,70	0,050	0,10	11	14	17	24
	0,025	0,05	13	17	20	27
0,75	0,050	0,10	10	12	14	20
	0,025	0,05	11	14	17	22
0,80	0,050	0,10	8	10	12	16
	0,025	0,05	10	12	14	18
0,85	0,050	0,10	7	8	10	13
	0,025	0,05	8	10	11	15

**Tabla J.** Número de sujetos necesarios para estimar una asociación entre dos variables cuantitativas mediante el coeficiente de correlación de Pearson (cont.)

r*	$\alpha$		Potencia ( $1 - \beta$ )			
	Hipótesis unilateral	Hipótesis bilateral	0,80	0,90	0,95	0,99
0,90	0,050	0,10	6	7	8	10
	0,025	0,05	7	8	9	11
0,95	0,050	0,10	5	6	6	8
	0,025	0,05	5	6	7	8

\*Valor estimado del coeficiente de correlación de Pearson que se desea detectar.

La fórmula para calcular el número de sujetos necesarios para estimar el coeficiente de correlación entre dos variables cuantitativas se encuentra en la tabla A7.1. El ejemplo A7.6 ilustra su uso.

**Tabla K.** Número de sujetos necesarios por grupo para detectar una determinada razón de riesgos en estudios que utilizan análisis de supervivencia

$\theta$	$\alpha = 0,05$ (bilateral)			$\alpha = 0,01$ (bilateral)		
	Potencia ( $1 - \beta$ )			Potencia ( $1 - \beta$ )		
	0,80	0,90	0,95	0,80	0,90	0,95
1,2	472	632	782	704	897	1.075
1,4	139	186	230	207	264	316
1,6	71	96	118	106	135	162
1,8	46	61	76	68	87	104
2,0	33	44	55	49	63	75
2,2	26	34	42	38	48	58
2,4	21	28	34	31	39	47
2,6	18	23	29	26	33	40
2,8	15	20	25	23	29	34
3,0	13	18	22	20	25	30
3,2	12	16	20	18	23	27
3,4	11	15	18	16	20	24
3,6	10	13	16	15	19	22
3,8	9	12	15	14	17	21
4,0	9	11	14	13	16	19
4,2	8	11	13	12	15	18
4,4	8	10	12	11	14	17
4,6	7	10	12	11	13	16
4,8	7	9	11	10	13	15
5,0	7	9	11	10	12	14

$\theta$ : cociente entre los riesgos de desarrollar el desenlace entre los grupos que se desea detectar. El ejemplo A7.8 ilustra su uso.

**Tabla L. Números aleatorios**

68 68 80 54 00	16 01 92 58 21	65 12 64 64 70	07 28 66 61 59	48 79 74 73 72	08 64 80 91 38
92 36 48 69 45	89 84 05 34 47	09 12 81 93 63	46 13 95 65 96	88 09 31 54 88	97 96 86 01 69
97 50 71 39 79	51 99 98 44 39	99 35 72 61 22	42 36 31 16 59	12 75 10 60 36	80 66 39 94 97
57 93 08 35 69	08 12 60 39 23	89 96 34 22 37	96 18 69 06 30	61 73 84 89 18	26 02 04 37 95
44 71 38 40 37	69 99 47 26 52	89 85 33 22 80	66 10 71 44 05	48 06 30 00 18	03 30 28 55 59
70 72 51 49 73	88 91 28 79 50	81 83 33 98 29	88 77 90 45 59	71 42 14 96 55	98 59 96 01 36
56 42 78 54 06	59 45 27 08 51	68 82 34 08 83	67 98 36 65 56	85 64 23 85 41	64 72 08 59 44
80 61 68 44 19	84 27 17 30 37	62 42 07 12 63	95 39 06 35 63	48 69 49 02 58	98 02 50 58 11
37 79 95 02 66	65 45 53 41 07	61 02 73 36 85	90 54 33 65 84	14 83 46 74 11	76 66 63 60 08
35 36 82 82 59	48 28 01 83 84	61 38 93 73 68	22 30 95 69 72	09 11 21 91 73	97 28 44 74 06
43 40 13 35 45	93 18 31 83 45	80 58 35 06 88	47 19 63 92 75	54 52 62 29 91	53 58 54 66 05
78 36 26 24 06	18 26 32 96 83	74 93 55 39 26	73 87 96 76 23	50 58 45 27 57	14 96 39 64 85
98 84 48 42 92	45 62 63 40 88	60 42 17 18 48	69 63 21 83 41	35 69 34 10 94	32 22 52 04 74
84 65 43 07 30	26 22 59 28 27	59 62 37 95 42	33 56 90 92 57	38 58 22 14 79	24 32 12 38 42
23 79 80 71 37	33 26 74 03 30	49 54 36 85 14	58 61 52 27 03	74 22 19 13 48	30 28 01 92 49
17 79 96 52 35	05 53 15 26 70	37 03 08 98 64	78 35 22 22 88	04 69 22 64 07	04 73 25 74 82
33 48 32 91 54	98 61 70 48 22	53 26 79 20 38	58 70 61 43 97	68 50 64 55 75	42 70 32 09 60
82 17 18 17 14	85 13 41 38 10	95 28 12 73 23	34 78 77 60 25	16 47 61 43 77	83 27 19 70 41
91 95 43 81 14	04 41 66 09 76	84 31 64 64 08	47 42 80 61 03	20 50 73 40 95	24 77 95 73 20
83 59 89 65 06	03 10 16 16 82 24	21 46 51 44 77	33 11 49 15 16	39 58 20 12 39	82 77 02 18 88
22 32 61 43 75	54 08 18 07 04	12 53 67 51 54	97 30 53 62 38	92 17 63 36 75	33 14 11 11 78
49 48 55 11 39	68 50 33 31 47	16 28 25 82 98	86 62 93 66 71	15 64 88 75 27	04 51 41 61 96
39 07 30 60 77	39 90 89 86 77	64 21 91 15 82	54 80 67 78 66	46 86 86 88 86	50 09 13 24 91
52 49 41 73 46	56 50 45 94 25	12 77 40 70 14	68 47 37 10 84	48 28 48 30 51	60 73 73 03 87
47 10 62 31 28	59 77 64 59 90	44 37 33 53 17	13 26 98 86 29	58 92 62 50 18	93 09 45 89 06
91 67 48 57 10	52 62 24 19 94	25 47 57 91 13	13 50 63 04 23	62 74 29 92 24	64 94 63 15 07
49 92 05 12 07	55 98 78 10 70	47 46 41 90 08	78 66 28 55 80	44 31 52 43 07	19 83 94 62 94
67 95 07 76 30	44 67 32 23 13	12 72 72 27 77	51 57 32 22 27	28 30 62 58 83	13 08 60 46 28
54 50 06 44 75	46 44 33 63 71	56 59 75 36 75	66 86 65 64 60	35 14 82 56 80	59 78 98 76 14
56 23 27 19 03	30 80 05 19 29	65 00 51 93 51	19 83 52 47 53	18 88 26 95 54	48 00 26 43 85
11 28 94 15 52	89 99 93 39 79	34 87 96 23 95	51 78 57 26 17	94 61 47 03 10	88 84 59 69 14
77 55 33 62 02	62 03 55 86 57	70 39 83 66 56	79 68 96 26 60	13 13 38 69 96	43 83 10 13 24
35 96 29 00 45	17 27 27 51 26	48 21 47 74 63	73 52 93 70 50	91 59 11 38 44	05 30 08 46 32
02 84 48 51 97	19 79 95 07 21	09 04 03 35 78	63 99 25 69 02	22 95 30 19 29	20 25 36 70 69
49 90 21 69 74	93 74 21 86 33	71 30 32 06 47	86 28 30 02 35	20 39 84 95 61	99 78 78 83 82
75 27 28 52 13	17 18 16 90 46	56 33 24 87 36	97 96 47 59 97	87 73 19 38 47	83 43 32 26 26
92 42 85 04 31	42 37 13 81 83	97 50 81 79 59	93 41 69 96 07	69 33 17 03 02	42 69 60 17 42
34 59 43 36 96	35 63 02 31 61	40 33 04 46 24	40 24 74 36 42	56 39 59 89 63	78 22 87 10 88
71 63 94 94 33	26 41 77 63 37	18 78 80 36 85	06 06 16 25 98	64 13 09 37 11	31 66 60 65 64
61 63 00 25 92	98 10 39 33 15	39 46 67 21 17	97 81 26 03 89	98 78 80 63 23	20 96 06 79 80
32 15 99 67 43	56 16 88 87 60	90 14 79 61 55	65 99 59 97 84	18 62 85 28 24	39 65 01 73 91
68 49 20 43 29	31 85 33 69 07	60 22 66 72 17	16 91 21 32 41	85 66 48 38 73	48 78 58 08 88
50 93 19 35 56	78 12 03 09 70	35 26 99 18 25	62 03 89 26 32	43 61 00 66 42	33 86 76 71 66
92 20 32 39 67	98 81 99 37 29	37 11 05 75 16	92 27 73 40 38	68 52 16 83 34	48 32 72 26 95
21 68 40 95 79	95 66 39 01 09	00 84 14 36 37	45 51 94 69 04	97 00 12 91 33	83 97 68 95 65
13 81 20 67 58	03 35 63 05 77	12 08 05 75 26	00 81 06 28 48	01 52 48 69 57	02 41 03 89 33
25 76 01 54 03	72 93 78 04 36	60 60 29 99 93	05 06 42 24 07	96 88 22 46 94	60 73 04 84 98
83 79 68 20 66	70 81 10 94 91	60 09 71 87 89	12 68 46 55 89	91 51 63 27 95	88 29 04 79 84
61 58 87 08 05	85 79 76 48 23	67 85 72 37 41	07 79 26 69 61	98 55 83 46 09	41 49 36 83 43
48 84 60 37 65	32 25 34 03 36	62 95 80 97 63	52 16 16 23 56	01 98 00 89 85	10 10 71 19 45

# Índice alfabético

## A

Adaptación transcultural, 207  
selección, 208  
Ajuste, 270, 281  
problemas, 284  
técnicas, 283  
Alfa de Cronbach, 202  
Análisis  
de casos válidos o según protocolo, 229  
de subgrupos, 214, 219  
estratificado, 283, 288  
multivariante, 262, 284, 288  
por intención de tratar, 43, 45, 228, 229  
por protocolo, 43  
Artículo original, 316  
agradecimientos, 323  
bibliografía, 321  
discusión, 321  
figuras, 320  
introducción, 316  
material y métodos, 317  
resultados, 319  
resumen, 322  
tablas, 319  
título, 322  
Asignación  
aleatoria, 38, 166, 282  
estratificada, 169  
por bloques, 168  
por grupos, 170  
simple, 168  
técnicas adaptativas, 170  
ventajas, 166  
de los factores de estudio, 30  
Autores, 311, 323  
Autoría, 23

## B

Buena práctica clínica, 233  
Búsqueda bibliográfica, 113

## C

Calidad de la evidencia, 294  
Causalidad, 292  
criterios, 294  
tipos de estudio, 292  
Código de Núremberg, 16  
Coeficiente de correlación de Pearson, 347  
intracoclase, 348  
Comités Éticos de Investigación Clínica, 20, 242  
Committee on Publication Ethics, 313  
Comparabilidad, 214  
Comparaciones múltiples, 261  
Comunicación, 240  
científica, 22, 309  
Concordancia, 345  
interobservador, 93  
intraobservador, 93  
Conflicto de intereses, 21, 23  
Confusión, 11, 217, 280  
Conocimiento, 4, 22  
Consentimiento informado, 18, 20, 37  
Consistencia, 295  
interna, 202  
CONSORT, 33, 42  
Contraste de hipótesis, 255.  
V. también *Prueba estadística.*  
Control de calidad de los datos, 239  
Corrección de Bonferroni, 262  
Cuaderno de recogida de datos, 235

Cuestionarios, 132, 185, 188  
encuestas  
administradas por encuestador, 189  
telefónicas, 190  
entrevistas personales, 189  
formato, 196  
orden de las preguntas, 195  
preguntas  
abiertas, 190  
cerradas, 190  
prueba piloto, 197  
recogida de datos, 237c  
redacción de las preguntas, 191  
selección de los ítems, 191  
validación, 200  
Cuidados habituales, 36  
Curvas ROC, 342  
área bajo la curva, 342

## D

Datos precodificados, 236  
Declaración de Helsinki, 16, 22  
Definición de las variables, 182  
Densidad de incidencia, 213  
Diferencia  
de incidencias, 301  
de riesgos, 211. V. también  
*Diferencia de incidencias.*  
Diseño, 29  
*Double dummy*, 36

## E

Editores, 312  
Efecto  
Hawthorne, 10  
placebo, 10  
*Effect size.* V. *Tamaño del efecto.*

## Índice alfabético

- Emparejamiento (*matching*), 79, 282  
Enmascaramiento, 39  
  doble ciego, 39  
  limitaciones, 40  
  simple ciego, 39  
tipos, 39  
triple ciego, 39  
ventajas, 39  
Ensayo  
  clínico  
    aleatorio, 33, 350  
      exclusiones preasignación, 223  
      modalidades de análisis, 229  
  cruzado, 57  
    desventajas, 59  
    efecto  
      período, 58  
      secuencia, 58  
    ventajas, 59  
  factorial, 56  
  secuencial, 54  
comunitario, 59  
pragmático, 46  
Equipo investigador, 111, 232  
Error  
  aleatorio, 7  
  estándar de la media (EEM), 254  
  sistématico, 8. V. también *Sesgo*.  
Escala(s)  
  de Guttman, 195  
  de Likert, 194  
  de medida, 182  
  visual analógica, 195  
Especificidad, 339, 341  
Estadística, 249, 253  
  descriptiva, 253  
  inferencial, 253  
Estrategia de análisis, 211  
  estudio  
    analítico, 214  
    descriptivo, 212c  
  sujetos a incluir, 221  
  variables que controlar, 216  
Estudio. V. *Objetivo del estudio*.  
ambispectivo, 30  
analítico, 29, 214  
  estrategia de análisis, 214  
antes-después, 61  
clínico aleatorio  
  estrategia de análisis, 43  
  finalización anticipada, 43  
  población, 37  
  seguimiento de los sujetos, 41  
de asociación cruzada, 87  
de casos y controles, 74  
  anidados en una cohorte, 84  
  definición de caso, 75  
  identificación de los casos, 75  
número  
  de controles por caso, 78  
  de grupos control, 78  
selección  
  de los casos, 74  
  de los controles, 76  
de cohortes, 64, 222  
  cohorte  
    dinámica, 68  
    fija, 68  
  prospectivo, 64  
  retrospectivo, 64  
  ventajas y limitaciones, 70  
de correlación ecológica, 95  
de equivalencia, 33, 36, 230  
de no inferioridad, 36, 44, 230  
de prevalencia, 87  
  no respuestas, 89  
de superioridad, 36  
descriptivo, 29, 212  
  estrategia de análisis, 212  
ecológico, 94  
experimental, 30, 33, 49  
  desarrollo clínico, 49  
finalidad, 29  
longitudinal, 29  
negativo, 265  
observacional, 30  
pragmático, 13  
prospectivo, 30  
retrospectivo, 30  
transversal, 29, 222  
Ética, 15  
  investigación y práctica clínica, 17  
  requisitos éticos en  
    investigación, 19  
    y publicación científica, 22  
Evaluación de un proyecto, 244  
Evidencia, calidad de la, 294  
Evolución natural, 11
- F**
- Factores de confusión, 11, 216, 251, 280  
  ajuste, 282  
  características, 280  
Falacia ecológica, 95  
Fiabilidad, 173, 200  
  concordancia, 175  
  evaluación de la, 175  
  fuentes
- de error, 202  
de variabilidad, 173  
interobservador, 201  
repetibilidad, 175  
test-retest, 201  
variación  
  causada por  
    el instrumento, 175  
    el observador, 175  
  individual, 174  
Finalidad del estudio, 29  
Formación del personal, 236  
Fracción atribuible, 302  
Fuerza de la asociación, 294
- G**
- Gestión del estudio, 232  
*Good clinical practice. V. Buena práctica clínica.*  
Gradiente biológico, 296  
Grupo control, 9, 11, 272
- H**
- Hipótesis  
  alternativa, 256  
  nula, 256
- I**
- Incidencia, 328, 350  
  acumulada, 213, 328  
  densidad, 213, 329, 333  
Índice kappa, 345, 361  
Informe Belmont, 16  
Interpretación de resultados, 247  
Intervalo de confianza, 45, 142, 254, 259  
  diferencia entre  
    dos medias, 260c  
    dos proporciones, 260c  
  estimación de un parámetro, 254, 256c  
Investigación, 3, 15  
  analítica, 6  
  clínica y epidemiológica, bases metodológicas, 7  
  descriptiva, 6  
  finalidades, 6  
  orientación, 6  
  proceso, 5  
Investigador principal, 233
- L**
- Lectores, 313

**M**

- Manual de procedimientos, 234  
*Matching.* V. Emparejamiento.  
 Metaanálisis, 97  
 Método científico, 3  
 Modificación del efecto, 216, 217, 287, 288  
 Muestra, 253  
 Muestreo, 155  
   aleatorio  
     estratificado, 157  
     simple, 156  
   en múltiples etapas, 158  
   no probabilístico, 159  
   probabilístico, 156  
   sistemático, 159  
   variaciones, 253  
   ventajas, 155

**N**

- No respuestas, 222, 225  
 Número de casos que es necesario tratar, 351

**O**

- Objetivo del estudio, 113  
   identificación del problema, 113  
   objetivo específico, 116  
   pertinencia, 115  
   preguntas secundarias, 119  
   viabilidad, 115  
 $Odds ratio$ , 215, 301, 333

**P**

- Paradoja de Simpson, 12  
*Peer-review.* V. Revisión de manuscritos.  
 Pérdidas de seguimiento, 223, 225, 274  
 Período de preinclusión, 37  
 Placebo, 20, 35, 299  
 Plausibilidad biológica, 295  
 Población de estudio, 135, 269, 303  
   análisis de subgrupos, 305  
   características generales, 137  
   comparabilidad, 136  
   criterios de selección, 139, 303  
   muestra, 135  
   niveles, 135  
   no respuesta, 140  
   pérdidas de seguimiento, 140  
   población diana, 135  
   procedencia, 139

- representatividad, 136  
 riesgo basal de la población, 305

- Potencia, 257  
   estadística, 265  
     análisis estadístico, 269  
     cálculo, 267  
     diseño del estudio, 268  
     interpretación de los resultados, 270

- Precisión, 7  
 Pregunta de investigación, 19

- Prevalencia, 328, 331, 340  
   de período, 213  
   de punto, 213

- Principio(s)  
   éticos, 16, 17  
     autonomía, 18  
     beneficencia, 18  
     justicia, 17  
     no maleficencia, 18  
   de incertidumbre, 34

- PRISMA, 97

- Proporción, 327

- Protocolo  
   de estudio, 358  
   de investigación, 107  
   esquema, 109

- Prueba  
   diagnóstica, 90  
   estadística, 217, 265, 269, 364  
     bilateral, 257  
     elección, 217  
     error  
        $\alpha$ , 257  
        $\beta$ , 257  
   hipótesis  
     alternativa, 256  
     nula, 256

- Kruskal-Wallis, 365  
 no paramétricas, 364  
 paramétricas, 364  
 $t$  de Student-Fisher, 364  
 técnicas estadísticas

- multivariantes, 366

- $U$  de Mann-Whitney, 364  
 unilateral, 257

- piloto, 240

- Publicación, 22  
   fragmentada, 24  
   redundante, 24  
   secundaria, 24

**R**

- Razón, 327, 341  
   de probabilidad, 341  
   de una prueba

- negativa, 341  
 positiva, 341

- Reducción absoluta de riesgo, 303, 350

- Registros, 186

- Regresión a la media, 10

- Relación temporal, 296

- Relevancia clínica, 258, 261f

- Repetibilidad, 201

- Representatividad, 213

- Restricción, 282

- Retiradas, 226

- Revisión

- de los datos, 211  
   de manuscritos, 22  
   por pares, 244  
   sistemática, 97, 353  
   análisis  
     de la heterogeneidad, 353  
     de sensibilidad, 355  
     de subgrupos, 356

- efectos

- aleatorios, 354  
   fijos, 354  
 estudios observacionales, 103  
 metaanálisis acumulativos, 355  
 metarregresión, 356  
 pacientes individuales, 103  
 tabla de evidencias, 353

- Revisores, 312

- peer-review*, 312

- Riesgo

- atribuible proporcional.  
     *V. Fracción atribuible.*  
   relativo, 215, 333, 351

**S**

- Secuencia temporal, 29

- Seguimiento

- de una cohorte, 90  
   del estudio, 238

- Sensibilidad, 339

- al cambio, 205

- Sesgo, 251, 272

- de Berkson, 79

- de detección, 79, 275

- de duración de la enfermedad, 52

- de incorporación, 92

- de información, 9, 275

- error de clasificación diferencial, 277

- no diferencial, 276

- de memoria, 81, 277

- de Neyman, 79

## Índice alfabético

- Sesgo (*Cont.*)  
de participación, 52  
de publicación, 101, 356  
*funnel plot*, 356  
de selección, 8, 272  
de sospecha diagnóstica, 93  
de verificación, 93  
del criterio de referencia  
    imperfecto, 92  
del entrevistador, 82  
del idioma inglés, 101  
en la localización de estudios,  
    101  
por adelanto en el diagnóstico, 52  
Significación estadística, 250, 256,  
    258f, 266  
Sobreajuste, 270, 287  
Solicitud de ayuda, 242  
    evaluación, 244  
    preparación, 242  
Supervivencia selectiva, 274
- T**
- Tamaño  
de la muestra, 142, 269, 361  
casos y controles, 363  
coeficiente de correlación, 362  
comparación  
    de dos medias, 150  
    de dos proporciones, 148  
concordancia, 361  
corrección para poblaciones  
    finitas, 144  
ensayos cruzados, 362  
equivalencia de dos  
    intervenciones, 150
- errores  $\alpha$  y  $\beta$ , 145  
estimación  
    de un riesgo relativo, 149  
    de una media, 144  
    de una *odds ratio*, 150  
    de una proporción, 143  
hipótesis  
    bilaterales, 145  
    unilaterales, 145  
índice kappa, 361  
intervalos de confianza, 151  
magnitud de la diferencia, 146  
nivel de confianza, 143  
no respuestas, 153  
precisión, 143  
pruebas  
    de contraste de hipótesis,  
        145  
    diagnósticas, 361  
supervivencia, 363  
variable de respuesta ordinal,  
    150  
    del efecto, 205  
Tasa, 327
- U**
- Usual care. V. Cuidados habituales.*  
Utilidad práctica, 251
- V**
- Validez, 9, 176, 203  
    científica, 20  
    concurrente, 204  
    convergente, 204  
    de constructo, 204
- de contenido, 203  
de criterio, 203  
discriminante, 204  
divergente, 204  
error  
    causado por  
        el individuo, 176  
        el instrumento, 177  
        el observador, 176  
diferencial, 178  
    no diferencial, 177  
evaluación, 177  
externa, 13  
fuentes de error, 206  
interna, 13, 292  
lógica, 203  
predictiva, 204  
Valor predictivo  
    negativo, 340  
    positivo, 340  
Variable(s)  
    combinadas, 129  
    desventajas, 129c  
    ventajas, 129c  
de respuesta, 36, 118, 121, 269,  
    299, 364  
    evaluación de la respuesta,  
        123  
intermedia, 300  
subrogada, 300  
definición de las, 182  
intermedia, 287  
modificadora del efecto, 216  
subrogada, 126  
    características, 126c  
Viabilidad, 200  
Voluntarios, 275

## CASO ESTUDIO. Ensayo clínico fisioterapia

A continuación se presenta un ejercicio práctico guiado de diseño e interpretación de un ensayo clínico basado en el siguiente artículo. Se trata de un ensayo clínico que compara la fisioterapia con el consejo en pacientes con lumbalgia.

Frost H, Lamb SE, Doll HA, Carver PT, Stewart-Brown S. Randomised controlled trial of physiotherapy compared with advice for low back pain. *BMJ*. 2004;329:708-13.

Disponible en:

[www.bmjjournals.org/cgi/reprint/329/7468/708](http://www.bmjjournals.org/cgi/reprint/329/7468/708)

Intente responder a las preguntas usted mismo antes de ver las respuestas propuestas. Las respuestas no son las únicas ni necesariamente las mejores.

### Resumen del artículo

**Objetivo.** Medir la efectividad de la fisioterapia de rutina comparada con una sesión de valoración y consejo por parte de un fisioterapeuta en pacientes con lumbalgia.

**Diseño.** Ensayo clínico aleatorio pragmático multicéntrico.

**Emplazamiento.** Siete departamentos de fisioterapia británicos.

**Participantes.** Un total de 286 pacientes con lumbalgia de más de 6 semanas de duración.

**Intervención.** Fisioterapia de rutina o consejo de un fisioterapeuta.

**Principales medidas de resultado.** La variable principal fue la puntuación del índice de discapacidad de Oswestry a los 12 meses. Las variables secundarias fueron las puntuaciones del índice de discapacidad de Oswestry a los 2 y 6 meses, del cuestionario de discapacidad de Roland y Morris y del SF-36 a los 2, 6 y 12 meses, así como el beneficio percibido por el paciente a los 2, 6 y 12 meses.

**Resultados.** De los 286 pacientes, 200 (70%) proporcionaron información de seguimiento a

los 12 meses. Los pacientes en el grupo de fisioterapia informaron mejores percepciones de beneficio, pero no se encontró evidencia de un efecto a largo plazo de la fisioterapia en las medidas de resultado genéricas ni en las específicas de la enfermedad (diferencia media en el índice de Oswestry a los 12 meses: -1,0%; IC 95%: -3,7% a 1,6%).

**Conclusiones.** La fisioterapia de rutina no parece ser más efectiva que una sesión de valoración y consejo por parte de un fisioterapeuta.

### EJERCICIO

La lumbalgia es la causa más frecuente de discapacidad física en la población en edad laboral. La fisioterapia es una intervención que habitualmente se centra en los factores físicos, incluyendo una combinación de movilización articular, consejo y programas individualizados de ejercicios.

En el Reino Unido, los fisioterapeutas tratan alrededor de 1,3 millones de pacientes por lumbalgia cada año, pero sólo existe una evidencia débil de la efectividad de la fisioterapia de rutina y ninguna de la efectividad de la electroterapia, del tratamiento con láser o ultrasonidos, o la tracción.

Las guías internacionales son variables, pero coinciden en aconsejar a los pacientes con lumbalgia que permanezcan físicamente activos, y prescribir los analgésicos apropiados. Recomiendan la terapia con ejercicios para pacientes con lumbalgia crónica (más de 12 semanas de duración) y algunas sugieren la manipulación espinal para la lumbalgia aguda o subaguda.

Los autores comentan que han investigado la efectividad de la fisioterapia, tal como se practica habitualmente en el Reino Unido, al cabo de un año, en pacientes con lumbalgia comparada con una sesión de valoración y consejo por parte de un fisioterapeuta.

### PREGUNTA 1

¿Cree que el objetivo está bien definido? Identifique los elementos que componen el objetivo de un estudio analítico.

Los elementos que componen el objetivo son:

- 1) **Pacientes/problema de salud:** pacientes con lumbalgia.
- 2) **Intervención de estudio:** fisioterapia de rutina.
- 3) **Intervención de comparación:** sesión de valoración y consejo realizada por un fisioterapeuta.
- 4) **Variable de resultado:** índice de discapacidad de Oswestry.

En la formulación del objetivo, tal como se cita en la introducción, no se especifica la variable de resultado, aunque puede identificarse a partir del resumen.

## Métodos de investigación clínica y epidemiológica

Los autores describen el método en los términos siguientes:

### Métodos

El protocolo del grupo de tratamiento permitió a los fisioterapeutas seleccionar entre un abanico de técnicas de terapia manual y de ejercicios. Los pacientes fueron reclutados en el entorno sanitario en que recibían el tratamiento.

### Participantes y elegibilidad

Se llevó a cabo un ensayo clínico aleatorio pragmático multicéntrico, enmascarado para el investigador, con consentimiento escrito de todos los pacientes.

Los participantes se identificaron a partir de las derivaciones a fisioterapia realizadas por los médicos generales, y fueron invitados a participar en 7 departamentos de fisioterapia.

Los terapeutas investigadores de cada centro contactaron con los pacientes para comprobar la elegibilidad. La valoración inicial incluía preguntas para descartar la posibilidad de que los participantes tuvieran enfermedades graves.

Los criterios de inclusión fueron: edad igual o superior a 18 años, lumbalgia de al menos 6 semanas de duración con o sin dolor en la pierna o signos neurológicos. Se excluyeron los pacientes con patología grave, incluyendo enfermedad reumatólogica sistémica, problemas ginecológicos, espondilitis anquilosante, tumores, infección, cirugía previa en la columna y tratamientos por problemas físicos en el mes anterior. También se excluyeron los pacientes derivados para programas intensivos de rehabilitación funcional.

### Medidas de resultado

La variable principal de resultado fue la puntuación del índice de discapacidad de Oswestry a los 12 meses: del 0% (ausencia de discapacidad) al 100% (discapacidad total o encamado).

Las variables secundarias son las puntuaciones del índice de Oswestry a los 2 y 6 meses, y las puntuaciones del cuestionario de discapacidad de Roland y Morris a los 2, 6 y 12 meses. Este cuestionario incluye 24 ítems relacionados con las funciones habitualmente afectadas por la lumbalgia: de 0 (mejor salud) a 24 (peor salud). La salud general se midió con el cuestionario SF-36 a los 2, 6 y 12 meses (las puntuaciones más altas indican mejor

salud). La percepción de beneficio por el paciente se midió en una escala de 0 (sin beneficio) a 10 (máximo beneficio), y en una escala dicotómica (se percibe beneficio o no se percibe beneficio).

Los datos de seguimiento fueron recogidos mediante un cuestionario postal por un investigador que desconocía el grupo al que había sido asignado el paciente. Los sujetos que no contestaban recibieron un recordatorio por correo. Y en los casos en que no se obtuvo respuesta, se les llamó por teléfono.

### Intervención

Los pacientes fueron asignados aleatoriamente a recibir consejo para permanecer activos (grupo de consejo), o bien consejo y fisioterapia estándar (grupo de tratamiento). Ambos grupos recibieron un libro con recomendaciones que fueron discutidas con el fisioterapeuta. El consejo se dirigía a la promoción del autocuidado y la modificación de las creencias y la conducta. La fisioterapia se inició en la semana posterior a la asignación.

#### Grupo de consejo

Los pacientes del grupo de consejo recibieron una sesión con un fisioterapeuta, que realizó un examen físico y proporcionó consejos generales para permanecer activo, de acuerdo con las recomendaciones del libro. La sesión tenía una duración aproximada de una hora.

#### Grupo de tratamiento

A los pacientes del grupo de tratamiento, los fisioterapeutas les realizaron un examen físico de una hora de duración. De acuerdo con la práctica habitual, los fisioterapeutas escogieron una estrategia terapéutica basándose en sus hallazgos, pero acordaron realizar el tratamiento de acuerdo con un protocolo estandarizado que refleja la práctica de rutina. Se incluía cualquier combinación de técnicas (movilización y manipulación articular, técnicas de tejidos blandos, ejercicios, tratamiento con frío o calor, consejo, etc.). El protocolo especificó hasta cinco sesiones adicionales de unos 30 minutos. Los fisioterapeutas registraron el tipo y el número de sesiones.

### Asignación y enmascaramiento

Un investigador determinó la secuencia de asignación antes del estudio utilizando una secuencia

de números aleatorios generada por ordenador. La asignación se codificó y guardó en sobres opacos, sellados y numerados secuencialmente. No se realizó ninguna estratificación.

La asignación se ocultó a los terapeutas que realizaban la valoración inicial, las personas que registraban los datos y el estadístico que realizó el análisis. El código era conocido exclusivamente por los fisioterapeutas que realizaban la interven-

ción. La asignación también se ocultó a los pacientes hasta la primera cita.

Aunque la asignación se ocultó a los investigadores, no fue posible enmascarar la intervención a los pacientes y a los fisioterapeutas más allá de la valoración inicial por motivos prácticos y éticos.

La valoración inicial fue realizada antes de la asignación.

## PREGUNTA 2

### **Los autores definen el ensayo como pragmático. ¿Qué significa este término?**

Significa que el estudio se ha diseñado intentando reproducir las características de la práctica habitual, de manera que busca obtener conclusiones que sean fácilmente generalizables y extrapolables a la práctica. Esta orientación del diseño se ve, por ejemplo, en el mecanismo de selección de los pacientes (sujetos derivados por médicos generales para fisioterapia) o en la intervención de estudio (los fisioterapeutas pueden elegir la técnica que deseen entre un amplio abanico de terapias manuales y de ejercicios).

El inconveniente principal de este enfoque es que se pierde capacidad de control sobre las condiciones del estudio, lo que puede limitar la validez interna de las conclusiones, ya que pueden existir factores no controlados que interfieran en los resultados.

## PREGUNTA 3

### **¿Cuáles son los principales criterios de selección de los sujetos que se incluyen en el estudio?**

Se incluyen los sujetos de 18 años o más con lumbalgia de al menos 6 semanas de duración, tengan o no dolor en la pierna o signos neurológicos. Se excluyen los que presentan problemas de salud importantes, los que han sido tratados por problemas físicos el mes anterior y los que han sido derivados para programas de rehabilitación funcional intensivos.

## PREGUNTA 4

### **¿Cree que el tipo de pacientes incluido es el adecuado para el estudio? ¿Cree que el estudio tiene validez externa?**

Los autores han realizado un enfoque pragmático del diseño, por lo que no han definido unos criterios de selección muy estrictos, sino que han intentado reproducir lo que ocurre en la práctica de su sistema sanitario. Tan sólo han excluido los sujetos que presentan problemas de salud importantes, que probablemente interferirían en la realización del estudio y que requerirían un abordaje diferente, y aquellos en los que la intervención de comparación supondría un retraso que podría causar un daño permanente o riesgo vital.

En cuanto a la capacidad de extrapolación a nuestro entorno, puede variar en función del tipo de centro en el que se trabaje o del tipo de pacientes a los que se atienda. Cuanto más parecidos son los sujetos que se atienden en un centro y los incluidos en el estudio, más fácil es extrapolar los resultados. Por ello es importante que los autores incluyan una descripción detallada de los pacientes estudiados. Probablemente, el procedimiento de derivación de los pacientes en el Reino Unido sea diferente del de nuestro país, lo que podría dificultar la extrapolación. Además, depende de si se trabaja en atención primaria, en una unidad de fisioterapia o rehabilitación, en un hospital, en una unidad de dolor, etc. El tipo de pacientes que se atiende en cada uno de estos contextos puede ser muy diferente (o no) del incluido en este estudio.

## PREGUNTA 5

### **¿Cuál es la variable de resultado principal?**

El instrumento de medida de la respuesta es el índice de discapacidad de Oswestry a los 12 meses, cuyos valores van del 0% (sin discapacidad) hasta el 100% (totalmente discapacitado o encamado).

## Métodos de investigación clínica y epidemiológica

La variable que se comparará es la diferencia de la puntuación de este índice entre la situación inicial (basal) y final (a los 12 meses) de cada sujeto.

### PREGUNTA 6

**¿Qué variables de resultado secundarias se miden? ¿Cuál es su utilidad?**

Los autores también deciden valorar otras variables además de la principal:

- El índice de discapacidad de Oswestry a los 2 y 6 meses.
- El cuestionario de Roland y Morris a los 2, 6 y 12 meses.
- El cuestionario de salud general SF-36 a los 2, 6 y 12 meses.
- El beneficio percibido por el paciente medido tanto en una escala numérica (de 0 a 10 puntos) como dicotómica (con o sin beneficio).

En principio, la utilidad de las variables secundarias es ayudar a interpretar los resultados del estudio al permitir comparar el efecto de las intervenciones sobre otros aspectos relevantes del problema de salud del estudio. Idealmente, deben reflejar todo el rango de resultados de interés, por lo que deberían permitir evaluar el efecto de la intervención sobre todas las variables importantes, incluyendo la seguridad y los efectos secundarios y adversos, especialmente cuando se estudian fármacos.

### PREGUNTA 7

**Una de las variables de resultado secundarias es el beneficio percibido por el paciente. Dado que se trata de una variable subjetiva, ¿cree que proporcionará información adicional útil y que es importante tenerla en cuenta?**

Cada vez se concede más importancia a las variables de resultado que incorporan la valoración que hacen los pacientes y sus preferencias. Este aspecto es especialmente importante en problemas tan subjetivos como puede ser el dolor, en que la experiencia del paciente es fundamental.

### PREGUNTA 8

**¿Cuáles son las intervenciones que se comparan?**

**Grupo de comparación:** sesión de una hora con un fisioterapeuta, durante la cual éste realiza una exploración física y da consejo al paciente para que permanezca activo, de modo que se promueva el autocuidado y se modifiquen las creencias y conductas. Se proporciona al paciente un libro con los consejos.

**Grupo de intervención:** tras una exploración física, el fisioterapeuta decide la estrategia terapéutica de acuerdo con la práctica habitual del NHS, y elige la que considera más adecuada entre diferentes combinaciones de técnicas. También se proporciona a los pacientes el mismo libro de consejos que en el otro grupo.

### PREGUNTA 9

**El hecho de que los pacientes del grupo de intervención no reciban todos el mismo tratamiento, ¿cree que invalida los resultados del estudio?**

No, ya que la intención de los autores es comparar la práctica habitual en estos pacientes. El problema es que hay un elevado componente de heterogeneidad en este grupo, lo que puede dificultar la interpretación y la obtención de resultados significativos. Los autores han intentado reducir este problema utilizando un protocolo estandarizado.

Deberían conocerse mejor los criterios de elección de cada estrategia de abordaje de los pacientes y hubiera sido interesante comparar subgrupos de sujetos en función de la estrategia terapéutica que han seguido, aunque sea de forma simplemente descriptiva y exploratoria.

Si se hubiera elegido una única estrategia y se hubiera realizado la misma en todos los sujetos, se estaría comparando esta estrategia frente al consejo aislado, lo que probablemente tendría mayor validez interna, pero hubiera limitado la generalización de los resultados a la práctica, al no proporcionar información sobre lo que se hace habitualmente en el abordaje de estos pacientes. Por ello, el enfoque del ensayo se define como pragmático.

Otra posibilidad sería haber utilizado diferentes grupos de intervención, de modo cada uno recibiera una estrategia terapéutica diferente. El inconveniente es que hubiera aumentado mucho el número de sujetos necesarios y hubiera hecho más compleja la realización y el análisis del estudio.

### PREGUNTA 10

#### ¿La asignación de los sujetos a los grupos puede considerarse realmente aleatoria?

Sí, dado que se realizó de acuerdo a una secuencia de números aleatorios generada por un ordenador.

### PREGUNTA 11

#### ¿La secuencia de asignación se ocultó a los investigadores? ¿Es importante ocultarla? ¿Por qué?

La secuencia se ocultó, ya que se utilizaron sobres cerrados y opacos, de manera que los investigadores desconocieran a qué grupo se asignaría un sujeto hasta haber decidido su inclusión.

La importancia de esta ocultación radica en evitar que el conocimiento de a qué grupo se asigna un sujeto pueda influir sobre la decisión de incluirlo o no. Así, por ejemplo, si un investigador supiera que un paciente se asignaría al grupo de consejo aislado, podría hacer que decidiera no incluirlo por considerar que preferiría realizar otra estrategia terapéutica. Si lo desconoce, debe valorar detalladamente si cumple los criterios de selección para que éstos sean los únicos criterios en los que base su decisión de incluir o no a un paciente.

### PREGUNTA 12

#### ¿Qué enmascaramiento se ha realizado en este estudio? ¿Por qué no es doble ciego?

Además de la ocultación de la secuencia de asignación, en este estudio también se enmascaró al investigador durante el examen clínico inicial (para evitar que interfiriera en la evaluación de la línea basal), a la persona que recoge los datos (para evitar que pueda interferir en los resultados de los cuestionarios al saber a qué grupo pertenece el sujeto) y al estadístico que analiza los datos (para evitar posibles manipulaciones del análisis).

No es doble ciego porque no es factible: si un grupo recibe solamente una sesión de una hora con consejos y el otro va acudiendo a sesiones de acuerdo con un plan terapéutico definido, ¿cómo enmascararlo? Lo único que se puede hacer es intentar que la persona que recoge los datos del estudio desconozca el grupo a que pertenece cada paciente (evaluador ciego), aunque incluso esto puede resultar difícil en la práctica, ya que pueden existir muchos detalles en la información que da el paciente que podrían indicar al evaluador si está realizando fisioterapia o no.

### PREGUNTA 13

**En relación con el tamaño de la muestra, los autores comentan que: «Se estimó a priori el tamaño de la muestra necesario, asumiendo una potencia del 90% y un valor alfa de 0,05. El número de sujetos se determinó para detectar una mínima diferencia clínicamente importante entre grupos de 4 puntos en la puntuación del índice de Oswestry, obteniéndose un tamaño de 112 pacientes en cada grupo (224 en total). Se decidió incluir al menos 270 pacientes para compensar pérdidas durante el seguimiento».**

**Los investigadores definen como la mínima diferencia de relevancia clínica entre los grupos una diferencia de 4 puntos en el índice de Oswestry. ¿Cómo interpreta este dato?**

Significa que considerarán que la fisioterapia es más efectiva si consigue, en promedio, que la disminución en la puntuación del índice de Oswestry sea como mínimo 4 puntos superior a la que se consigue solamente con el consejo.

Lo que debemos preguntarnos es si una diferencia de 4 puntos en este índice (en una escala de 0 a 100) tiene o no relevancia clínica. De todas maneras, nos indica que éste es el criterio de los investigadores.

### PREGUNTA 14

**Los investigadores deciden utilizar una potencia estadística del 90%. ¿Le parece suficiente?**

Significa que, si en realidad la diferencia entre los grupos en la disminución del índice de Oswestry es de 4 puntos o más, el estudio tendría un 90% de posibilidades de detectarla con el número de sujetos previsto.

Una potencia del 90% es suficiente, ya que habitualmente se recomienda que sea, como mínimo, del 80%.

### PREGUNTA 15

**Los autores comentan que, de los 508 pacientes en los que evalúan la elegibilidad, incluyen 286 (56,3%), y asignan 144 al grupo tratamiento y 142 al grupo consejo. ¿Le parece coherente con el cálculo del tamaño de la muestra que han realizado?**

Se asignan finalmente 286 pacientes, 142 al grupo control y 144 al de intervención.

Dado que el cálculo del tamaño de la muestra indica el mínimo número de sujetos necesario en función de determinadas asunciones, el hecho de incluir un número ligeramente superior es adecuado.

Por otro lado, solamente el 56% de los candidatos potenciales son incluidos en el estudio. Es importante conocer las características y los motivos por los que los restantes pacientes no han sido incluidos, ya que pueden suponer una limitación para la generalización (validez externa) de los resultados.

### PREGUNTA 16

**En relación con la comparabilidad inicial de los grupos, los autores comentan que el grupo de consejo tenía una proporción ligeramente superior de varones y de fumadores, y que ambos grupos estaban equilibrados por el resto de las variables.**

**Supongamos que, antes del inicio del estudio, los investigadores tuvieron la hipótesis de que la fisioterapia tiene resultados diferentes en los varones y en las mujeres. En este caso no interesa que exista una diferente distribución de varones y mujeres en los grupos del estudio. ¿Qué hubieran podido hacer en el diseño para evitar que se produjeran estas diferencias?**

Para conseguir que ambos grupos de estudio tuvieran una distribución similar de sujetos en función del sexo, hubiera podido realizarse una asignación aleatoria estratificada por sexo, de manera que hubiera quedado la misma proporción de varones y mujeres en ambos grupos.

Los investigadores dicen que no estratificaron la asignación, pero al presentar la estrategia de análisis informan de que decidieron utilizar una técnica de análisis ajustando por esta variable (y otras) para neutralizar su posible efecto sobre los resultados.

### PREGUNTA 17

**Finalizan el estudio 97 pacientes del grupo de consejo y 103 del grupo de tratamiento. ¿Cómo interpreta este dato teniendo en cuenta el resultado del cálculo del número de sujetos necesario? ¿Cree que el porcentaje de pérdidas es excesivamente elevado y puede invalidar el estudio?**

Se habían asignado 286 pacientes, 142 al grupo consejo y 144 al tratamiento. Al final del estudio, al año de seguimiento, quedan 97 y 103, respectivamente, lo que supone unas pérdidas de 45 (31,7%) y 41 (28,5%) sujetos, respectivamente.

El porcentaje de pérdidas observado es ligeramente superior al previsto por los investigadores al calcular el tamaño de la muestra, de manera que finalizan el estudio 200 sujetos, menos de los 224 necesarios según su cálculo, lo que implica que la potencia real del estudio ha sido (ligeramente) inferior a la deseada.

Un porcentaje de pérdidas del 30% puede considerarse elevado (algunos autores consideran que el máximo aceptable es del 20%). Aunque ello no invalida necesariamente los resultados del estudio, supone una limitación muy importante que obliga a adoptar precauciones especiales al interpretarlos, sobre todo en los estudios con resultados estadísticamente no significativos, como en este caso.

En realidad, el impacto de las pérdidas sobre los resultados depende fundamentalmente de los motivos por los que se han producido, y por ello los autores del artículo deberían informarnos de los motivos de las pérdidas y de cómo podrían haber influido sobre los resultados, especialmente dado su elevado número.

Los autores compararon las características de los sujetos que completaron el estudio con las de los que no lo hicieron, y dicen que sólo observaron pequeñas diferencias en el sentido de que los que finalizaron el estudio eran algo mayores en edad, menos fumadores y con menor probabilidad de tener un primer episodio de lumbalgia o una historia de lumbalgia crónica. Los investigadores consideran que estas diferencias son pequeñas y tienen un impacto reducido sobre los resultados.

### PREGUNTA 18

**Para el análisis de los sujetos que no completan el seguimiento, los investigadores sustituyen los datos ausentes (*missing*) con el último valor disponible que tenían de los sujetos (*last known value carried*). ¿Cree que esto puede interferir en los resultados?**

Esta estrategia de análisis consiste en sustituir el dato que falta de un sujeto con el último de dicho sujeto que se tenga disponible de las visitas anteriores. Se asume, así, que tendría el mismo resultado que en la visita anterior. Si ello no es así, se puede introducir un sesgo en el análisis. Por ejemplo, si las diferencias en el resultado del tratamiento se producen más tarde, el sustituir los valores finales por los que se tienen en visitas más tempranas puede conducir a que no se pongan de manifiesto dichas diferencias. Si las diferencias se producen al inicio pero se diluyen al final, la sustitución puede conducir a que aparentemente las diferencias sean mayores. Por lo tanto, es importante que, cuando se realiza esta sustitución de valores, los investigadores evalúen detalladamente su posible impacto sobre los resultados.

Los autores han analizado los resultados con y sin esta sustitución de valores, y nos dicen que no se producen cambios sustanciales, de modo que podemos suponer que en este caso no se ha introducido un sesgo importante. Sin embargo, no debe olvidarse que se ha producido en un importante número de casos (las pérdidas son de un 30%), lo que supone una limitación para la interpretación de los resultados.

### PREGUNTA 19

**Los autores dicen que han utilizado la estrategia de análisis por intención de tratar. ¿Cree que es la adecuada para este estudio?**

La estrategia de análisis por intención de tratar (o según asignación aleatoria) significa que cada sujeto se considera en el análisis en el grupo al que ha sido asignado, independientemente de si ha cumplido con la intervención asignada o si se ha perdido durante el seguimiento.

Esta estrategia de análisis es la más adecuada para un ensayo clínico que busca determinar si una intervención es superior a otra, ya que es la más conservadora y la que más dificulta encontrar resultados favorables a la intervención de estudio.

Sin embargo, debe tenerse en cuenta que, en caso de no encontrar diferencias estadísticamente significativas, se concluirá que no se ha demostrado superioridad, pero no podrá concluirse la equivalencia entre ambas intervenciones. Para ello hubiera sido necesario plantear el diseño del estudio como un estudio de equivalencia.

### PREGUNTA 20

**La variable de respuesta principal es la diferencia en la puntuación del índice de Oswestry a los 12 meses en relación con la línea basal. En la tabla 2 del artículo puede verse que el grupo de tratamiento ha disminuido en 3,27 puntos, y el grupo de consejo en 2,23 puntos. La diferencia entre ellos es de -1,04 puntos, con un intervalo de confianza del 95% cuyos límites son -3,70 y 1,59 puntos.**

**¿Cómo se interpreta este resultado? ¿La diferencia es estadísticamente significativa?**

La diferencia es de 1,04 puntos a favor del grupo de tratamiento, es decir, los sujetos que han recibido fisioterapia han disminuido en promedio 1,04 puntos más que el grupo control en el índice de Oswestry.

## Métodos de investigación clínica y epidemiológica

Sin embargo, el intervalo de confianza del 95% de este valor nos indica que el valor 0 queda incluido entre sus límites, por lo que el resultado no es estadísticamente significativo (es decir, el valor de p es superior a 0,05).

Por lo tanto, la interpretación de este dato es que no puede asegurarse que la fisioterapia produzca un descenso de las puntuaciones del índice de Oswestry superior al consejo solo. Ello no significa que la eficacia de ambos tratamientos sea la misma, sino simplemente que no se ha podido demostrar que exista una diferencia entre ellos.

### PREGUNTA 21

**¿Cree que la diferencia de la variable de respuesta entre los grupos de estudio es clínicamente relevante?**

La diferencia es de 1,04 puntos a favor del grupo de tratamiento, es decir, los sujetos que han recibido fisioterapia han disminuido en promedio 1,04 puntos más que el grupo control en el índice de Oswestry, aunque esta diferencia no es estadísticamente significativa.

Por otro lado, los límites del intervalo de confianza son -3,70 y 1,59, lo que indica que la verdadera diferencia entre los grupos está entre estos límites. Sería interesante conocer la sensibilidad al cambio del cuestionario, de manera que pueda determinarse hasta qué punto una disminución de 2 o 3 puntos en el índice de Oswestry supone una mejoría clínica real, o una diferencia de 1 punto entre dos grupos indica la existencia de una diferencia clínicamente relevante de eficacia.

De todas formas, los investigadores definieron en el cálculo del tamaño de la muestra que la mínima diferencia que consideraban de relevancia clínica era de 4 puntos. Como este valor queda fuera del intervalo de confianza, podemos asegurar que, aunque existiera una diferencia real entre los grupos, difícilmente ésta sería superior a 4 puntos, por lo que podemos descartar que exista una diferencia de relevancia clínica según el criterio de los investigadores.

### PREGUNTA 22

**En la tabla 4 del artículo puede verse que los pacientes del grupo de tratamiento tienen una mayor percepción de que se han beneficiado que los del grupo de consejo, tanto a los 2 meses (el 76,9 frente al 59,8%) como a los 6 meses (el 64,6 frente al 50,5%), como a los 12 meses en las puntuaciones de la escala de 0 a 10 puntos (5,02 frente a 4,13 puntos). Todas estas diferencias son estadísticamente significativas, con valores de p inferiores a 0,05.**

**¿Puede calcular el riesgo relativo de percepción del beneficio a los dos meses?**

El riesgo relativo de percepción del beneficio a los 2 meses del grupo de tratamiento en relación con el grupo de consejo puede calcularse dividiendo los porcentajes observados en ambos grupos, es decir,  $76,9/59,8 = 1,29$ . Este valor significa que el porcentaje de sujetos que consideran que se han beneficiado en el grupo tratamiento es 1,3 veces superior al del grupo consejo.

Los autores informan que los límites del intervalo de confianza del 95% de este riesgo relativo son 1,07 y 1,54. Dado que el intervalo excluye el valor 1, indica que este riesgo relativo es estadísticamente significativo, por lo que podemos asegurar que existe una mayor percepción de beneficio en el grupo de tratamiento, que es aproximadamente un 30% superior al del grupo de consejo.

### PREGUNTA 23

**¿Puede calcular cuántos sujetos sería necesario tratar con fisioterapia para conseguir que haya uno más que perciba a los 2 meses que se ha beneficiado?**

La diferencia en el porcentaje de sujetos que perciben beneficio a los 2 meses es de  $76,9 - 59,8 = 17,1$ . Por tanto, por cada 100 pacientes que reciban fisioterapia, habrá aproximadamente 17 sujetos más que perciban beneficio que si hubieran recibido solamente consejo. Es decir, el número de pacientes que es necesario tratar con fisioterapia para conseguir que haya uno más que perciba que se ha beneficiado que si hubiera recibido solamente consejo es aproximadamente de  $6(100/17)$ .

**PREGUNTA 24****A partir de los resultados comentados, ¿podría concluir que la fisioterapia no es efectiva?**

No. En primer lugar, el estudio ha sido diseñado como un estudio de superioridad, por lo que, en principio, solamente puede concluirse que no se ha encontrado que la fisioterapia sea superior al consejo aislado, pero tampoco que ambos sean iguales.

Además, el hecho de que se haya diseñado como un estudio pragmático en que el grupo de intervención ha recibido estrategias terapéuticas heterogéneas puede diluir el efecto, de manera que alguna(s) estrategia(s) sí sea(n) efectiva(s), pero al quedar diluidas en el conjunto, puede no haberse detectado.

Por otro lado, el grupo de comparación también recibía alguna intervención (consejo por parte de un fisioterapeuta además de un libro con recomendaciones para el autocuidado de su problema de salud). Para poder contrastar si la fisioterapia es efectiva o no, hubiera sido necesario un grupo control sin ninguna intervención, aunque probablemente esta comparación no sería factible, ya que podría ser considerada poco ética.

## CASO ESTUDIO. Estudio de prevalencia

A continuación se presenta un ejercicio práctico guiado de diseño e interpretación de un estudio descriptivo basado en el siguiente artículo:

Marín GH, Fazio P, Rubbo S, Baistrocchi A, Sager G, Gelembur A. *Prevalencia de anemia del embarazo y análisis de sus factores condicionantes*. Aten Primaria. 2002;29(3):158-63.

Intente responder a las preguntas usted mismo, antes de ver las respuestas propuestas. Las respuestas no son las únicas, ni necesariamente las mejores.

### EJERCICIO

La anemia es un problema frecuente durante el embarazo, y supone un incremento del riesgo tanto para la madre como para el hijo.

La valoración de la anemia debe tener en cuenta que, durante el embarazo, se produce un aumento

del volumen plasmático y de la masa eritrocitaria del 40 y el 25%, respectivamente. Este desequilibrio hace que, aun cuando la masa eritrocitaria aumente, pueda producirse una anemia por dilución. Además, las variaciones fisiológicas que se llevan a cabo durante un embarazo normal provocan un descenso en los niveles de hemoglobina (Hb) que no siempre constituye una anemia. Por esta razón, es incorrecto aplicar a las embarazadas los valores de Hb estipulados como límites para el diagnóstico de anemia en una mujer adulta no grávida. De hecho, el valor de  $Hb < 11\text{ g/dl}$  es el utilizado de forma más habitual para establecer este diagnóstico en las mujeres embarazadas.

Muy frecuentemente, las anemias durante el embarazo se originan en un déficit de nutrientes como hierro o ácido fólico, que suele asociarse a poblaciones de bajos recursos económicos y a embarazadas adolescentes y con carencias nutricionales.

Existe una amplia variabilidad geográfica en la frecuencia de este problema de salud.

### PREGUNTA 1

**Suponga que desea conocer la prevalencia de anemia en las mujeres embarazadas de una comunidad. ¿Qué tipo de estudio realizaría?**

Se trata de conocer la frecuencia de un problema de salud, por lo que el estudio corresponde a un diseño descriptivo. Dado que lo que se pretende determinar es la prevalencia, el diseño más adecuado es el de un estudio transversal (denominado precisamente de prevalencia).

### PREGUNTA 2

**Los autores se plantean un estudio para conocer la prevalencia de anemia en las embarazadas de la provincia de Buenos Aires (Argentina), dado que no existían datos publicados sobre el tema. ¿Cree que el objetivo está suficientemente definido?**

Sí. El objetivo de un estudio descriptivo debe incluir:

- Fenómeno que describir: prevalencia.
- Problema de salud: anemia.
- Población: mujeres embarazadas de la provincia de Buenos Aires.

### PREGUNTA 3

**En relación con las participantes en el estudio, los autores dicen que: «Se realizó un estudio prospectivo tomando como población de interés a un grupo de gestantes que consultaron en la región Capital, provincia de Buenos Aires, en el período enero-julio de 2000. Se incluyó a aquellas mujeres que consultaron por primera vez en dicha gestación, y como criterios de exclusión se consideraron la presencia de patología previa o la toma regular de fármacos o suplemento dietético».**

**¿Cuál es la población de estudio? ¿Cree que los criterios de selección están bien definidos?**

La población de estudio está definida por los siguientes criterios:

- Criterio de inclusión: mujeres embarazadas que consultan por primera vez durante una gestación en la región Capital, provincia de Buenos Aires, en el período enero-julio de 2000.
- Criterios de exclusión: presencia de patología previa o toma regular de fármacos o suplemento dietético.

La definición de los criterios de exclusión podría ser más clara. Así, debería indicarse con más detalle qué patologías son excluidas, o qué fármacos y suplementos dietéticos se considerarán criterios de exclusión, así como si se refiere a que los toman solamente durante el embarazo o a que su consumo es habitual, por ejemplo.

Por otro lado, en el criterio de inclusión, debería quedar más claro en qué centro o centros se incluyeron las pacientes (¿la región Capital es representativa de la provincia de Buenos Aires?), para poder evaluar la representatividad de las mujeres incluidas en relación con la población diana.

#### PREGUNTA 4

##### **¿Cuál es la población diana? ¿Cree que la población de estudio es representativa de la población diana?**

Tal como se define en el objetivo, la población diana está formada por las mujeres embarazadas de la provincia de Buenos Aires.

La representatividad se ve limitada por dos factores principalmente:

- Los criterios de exclusión. Si existe un elevado porcentaje de mujeres embarazadas que tienen patología previa, o que toman fármacos o suplementos dietéticos, y si estos factores pueden estar relacionados con la frecuencia de la anemia, entonces la población de estudio puede obtener un resultado incorrecto. Ello no significaría que la población de estudio estaría mal definida y se habría introducido un sesgo, sino simplemente que el resultado del estudio se refiere a la prevalencia en la población de estudio, y que este resultado no podría extrapolarse automáticamente a la población diana (limitación de la validez externa).
- Los criterios de inclusión. Algo parecido ocurre con el centro en que consultan las mujeres. Si las pacientes que acuden a estos centros no son representativas de las mujeres embarazadas de la provincia, o si la región Capital no es representativa de toda la provincia, entonces la validez externa puede verse comprometida.

#### PREGUNTA 5

##### **Los autores afirman que el riesgo de anemia se sitúa entre un 25 y un 65%. Si los autores esperaran encontrar un 25% de anemias, ¿cuántos sujetos hubiera sido necesario estudiar?**

Para determinar el tamaño de la muestra, es necesario definir:

- El valor esperado del parámetro de estudio. En este caso, la prevalencia de anemia, que según el enunciado de la pregunta se estima en un 25%.
- El nivel de confianza. Habitualmente es del 95%.
- La precisión de la estimación. Debe definirse la amplitud deseada del intervalo de confianza alrededor del valor esperado. En función de este valor, y consultando la tabla de determinación del tamaño de la muestra para la estimación de una proporción, serían necesarios:

Precisión	n
±5%	288
±4%	450
±3%	800
±2%	1.801
±1%	7.203

Por lo tanto, dependiendo de la precisión deseada por los investigadores, el tamaño necesario puede variar ampliamente. Lo importante es establecer una precisión que permita a los investigadores alcanzar su objetivo y obtener un resultado que les sea útil. El porcentaje de no respuestas esperado.

- En caso de que puedan existir no respuestas, debería aumentarse el tamaño de la muestra. Si, por ejemplo, se administra un cuestionario o debe solicitarse el consentimiento para el estudio, es posible que algunas mujeres no acepten participar o no contesten el cuestionario.

En el artículo, los autores no indican si han determinado el tamaño necesario ni si han tenido en cuenta la posibilidad de las no respuestas. Ello no significa que el estudio no sea válido, pero es conveniente determinar el número de sujetos necesario antes de iniciar el estudio.

### PREGUNTA 6

**El hecho de que la muestra no se haya seleccionado aleatoriamente, ¿compromete la validez del estudio?**

La muestra parece estar formada por todas las mujeres embarazadas que consultan durante un período de tiempo determinado. Por lo tanto, se han incluido todas las mujeres que cumplen los criterios de selección. La finalidad de muestreo es obtener una muestra representativa de la población, lo que también se logra incluyendo todos los sujetos. El problema es que puede haberse incluido un número innecesariamente elevado de mujeres.

La validez del estudio solamente se vería comprometida si se sospechara que la muestra incluida no es representativa de la población de estudio.

### PREGUNTA 7

**¿Hubiera podido aplicarse algún método de muestreo para seleccionar la muestra?**

Sí. Dependiendo del tamaño de la muestra necesario y del número de mujeres que consultan en los centros, en lugar de limitar el período de inclusión a 7 meses, se hubiera podido optar, por ejemplo, por realizar un muestreo sistemático, seleccionando, por ejemplo, una de cada 3 o 4 mujeres que consultan. Esta estrategia habría sido útil si el tiempo necesario para estudiar a cada mujer hubiera sido superior al habitual de una consulta, de manera que pudiera haberse recogido la información con mayor detalle sin restricciones de tiempo.

Si se sospechara que puede haber diferencias en la frecuencia de anemia según los meses de año, por ejemplo, hubiera podido plantearse un muestreo en etapas, seleccionando, por ejemplo, unos días aleatoriamente (o estratificadamente según el mes o la estación), y posteriormente estudiar todas (o una muestra) de las mujeres atendidas cada uno de los días seleccionados.

Las utilidades del muestreo podrían ser obtener una muestra representativa según diferentes factores (meses del año, centros que participan, tipo y profesional que atiende a la paciente, etc.), y ajustar el tamaño de la muestra al necesario, evitando la inclusión de un número innecesariamente elevado de mujeres, lo que es fundamental cuando el esfuerzo para la recogida de datos es importante.

### PREGUNTA 8

**En relación con la recogida de datos, los autores dicen: «Se diseñaron dos tipos de encuestas a realizar entre las embarazadas en forma prospectiva. La primera contempló la obtención de datos generales: edad, semanas de gestación, peso, talla, índice de masa corporal, ingreso familiar promedio (expresado en dólares), tipo de vivienda, acceso a agua potable y cloacas, patologías previas, antecedentes personales, número de gestaciones previas y nivel inicial de hemoglobina (expresada en g/dl).**

**El segundo elemento de obtención de datos fue una encuesta nutricional prospectiva, en la cual se hiciese constar de forma detallada toda la ingesta diaria durante el plazo de una semana. Para su mejor cumplimiento, se realizó a las gestantes una entrevista personal con un equipo de licenciadas en nutrición que utilizaron ejemplos prácticos y solución de dudas, destinados a enseñar la manera correcta de completar la encuesta.**

**A cada paciente de este estudio se le practicó una analítica inicial en la que se incluyeron parámetros hematológicos como hematocrito, hemoglobina, cifra de glóbulos rojos, volumen corpuscular medio, hemoglobina corpuscular media, concentración de hemoglobina corpuscular media, frotis sanguíneo para evaluar el aspecto de los glóbulos rojos, número de leucocitos, fórmula leucocitaria y cifra de plaquetas».**

**¿Cree que la definición de la variable principal es suficiente?**

La variable principal del estudio y la que debe definirse con la mayor precisión es la anemia. Los autores deberían haber establecido con claridad qué definen como anemia y garantizar la calidad del diagnóstico. Lo único que sabemos es lo que comentan en la introducción sobre los niveles de hemoglobina ( $Hb < 11 \text{ g/dl}$ ), pero hubiera sido preferible definirlo explícitamente en el apartado de «Material y métodos».

**PREGUNTA 9**

**El hecho de que la analítica se lleve a cabo en la primera visita que realiza cada paciente, ¿le parece correcto?**

Los autores comentan que puede existir un sesgo en los datos debido a la diversidad en el período en el cual se realizó la primera consulta (la mayoría de las pacientes lo hizo a finales del primer trimestre o promediando el segundo trimestre de embarazo), que puede considerarse tardía y presentar por tanto una mayor hemodilución.

Sin embargo, los autores también comentan que el sesgo no es importante, ya que hay evidencias que sostienen que a la mitad del embarazo el riesgo de presentar anemia es también de la mitad si se compara con el riesgo al final.

**PREGUNTA 10**

**Los autores informan que durante el período de estudio se recibieron 3.592 consultas a causa de embarazo, realizadas por 1.218 embarazadas que cumplieron los criterios de inclusión y no presentaban ningún criterio de exclusión. En 196 de las gestantes se detectó anemia.**

**La primera etapa de la estrategia de análisis es la descripción de la población estudiada. ¿Qué variables cree que es importante utilizar en dicha descripción?**

La finalidad de esta descripción es presentar las características de las mujeres estudiadas, de manera que los lectores puedan conocerlas y evaluar si son similares o no a las que atienden en su práctica.

Los autores presentan los siguientes valores:

- Edad:  $26,2 \pm 6,0$
- Talla:  $1,58 \pm 7,0$
- Peso en la primera consulta:  $63,96 \pm 20,40$
- Número de gestaciones previas:  $2,57 \pm 2,16$
- Semanas de gestación (primera consulta):  $21,04 \pm 11,21$

Así como los parámetros hematológicos y dietéticos.

**PREGUNTA 11**

**¿Cuál es la prevalencia de anemia detectada?**

$$\text{prevalencia} = \frac{196}{1.218} = 16\%$$

**PREGUNTA 12**

**¿Puede calcular el intervalo de confianza (IC) del resultado?**

La fórmula para el cálculo del IC es:

$$P \pm (Z_{\alpha} \times EEP)$$

siendo EEP el error estándar de la proporción:

$$EEP = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0,16(1-0,16)}{1.218}} = 0,01$$

Por lo tanto, el IC del 95% es:

$$\begin{aligned} \text{IC 95\%: } & 0,16 \pm 1,96 \times 0,01 \\ & 0,16 \pm 0,02 \end{aligned}$$

Los límites del IC son 0,14 y 0,18, es decir, del 14 y el 18%.

**PREGUNTA 13**

**¿Cómo interpreta el resultado?**

Existe un 95% de confianza de que la verdadera prevalencia de anemia en las mujeres embarazadas de la provincia de Buenos Aires, asumiendo que la muestra es representativa de la misma, está entre el 14 y el 18%.

### PREGUNTA 14

#### ¿Cree que un análisis de subgrupos podría aportar información de interés?

Un análisis de subgrupos permitiría comparar la prevalencia de anemia en las embarazadas según diferentes características (por ejemplo, nivel socioeconómico, número de gestaciones previas, etc.), lo que permitiría conocer mejor la frecuencia de este problema.

Los autores presentan algunos análisis adicionales, comparando las características de las embarazadas con y sin anemia. Concluyen que la anemia es más frecuente en mujeres con un peso inferior a 49 kg en la primera consulta de su gestación, en las que realizaron dicha consulta con posterioridad a las 20 semanas de embarazo o en las gestantes con ingresos familiares reducidos ( $<400$  dólares). También encontraron un mayor riesgo de anemia en las pacientes cuya dieta tenía un nivel proteíco  $<13\%$ , en las que consumían menos de 7 mg de hierro semanal o un porcentaje  $<10\%$  de hierro hémico.

También presentan algunas características de las anemias detectadas. Así, por ejemplo, informan que el 67% de las embarazadas anémicas tenía microcitosis e hipocromía, lo que se relaciona con un proceso de déficit de hierro.

Todos estos datos ayudan a matizar los resultados e interpretarlos correctamente.

## CASO ESTUDIO. Estudio de validación de un cuestionario

A continuación se presenta un ejercicio práctico guiado de diseño e interpretación de un estudio descriptivo basado en el siguiente artículo:

*Ruiz de Viñaspre Hernández R, Tomás Aznar C, Rubio Aranda E. Validación de la versión española de las formas cortas del Urogenital Distress Inventory (UDI-6) y del Incontinence Impact Questionnaire (IIQ-7) en mujeres embarazadas. Gac Sanit. 2011;25(5):379-84.*

Intente responder a las preguntas usted mismo, antes de ver las respuestas propuestas. Las respuestas no son las únicas, ni necesariamente las mejores.

### EJERCICIO

La incontinencia urinaria se define como cualquier pérdida involuntaria de orina. Su prevalencia en la población general de mujeres es alta, de entre el 20 y el 50%, y durante la gestación es aún mayor, con cifras por encima del 50%. Su impacto debe valorarse, además de por la presencia de síntomas y su intensidad, por el grado en que afecta a la calidad de vida de quien la padece.

Los cuestionarios elaborados para medir la percepción de la calidad de vida en las mujeres con incontinencia urinaria son autocumplimentados, lo que conlleva el riesgo de que sean interpretados

y contestados de manera diferente en diferentes poblaciones y culturas, por lo que se recomienda que sean probados y validados en la población en la que van a realizarse los estudios clínicos o epidemiológicos.

En la década de 1990 se desarrollaron los cuestionarios Incontinence Impact Questionnaire (IIQ) y Urogenital Distress Inventory (UDI), pero resultaban poco prácticos por su extensión. Posteriormente se desarrollaron sus versiones cortas, IIQ-7 y UDI-6, que demostraron un alto grado de correlación con sus respectivas versiones originales.

El UDI-6 mide los siguientes parámetros: presencia, gravedad, sintomatología urogenital asociada y tipo de incontinencia urinaria. Se divide en tres subescalas, en las cuales se agrupan los síntomas que recoge el cuestionario: subescalas de síntomas irritativos, de síntomas de estrés y de síntomas de obstrucción o dolor relacionados con la micción.

El IIQ-7 mide siete componentes de la calidad de vida en las mujeres con incontinencia urinaria: capacidad para hacer las tareas del hogar, actividad física, actividad recreativa, capacidad para viajar, actividades sociales, estado emocional y frustración.

Ambos cuestionarios se han utilizado para evaluar la incontinencia en las mujeres embarazadas, pero cuando fueron validados se excluyó del estudio a las gestantes.

### PREGUNTA 1

Dado que estos cuestionarios han sido validados en población femenina, ¿cree que es necesario validarlos también en mujeres embarazadas?

Antes de poder ser utilizado, un cuestionario debe ser validado en una población de sujetos de características similares a los que van a ser estudiados. Los autores informan de que los cuestionarios habían sido validados en mujeres no gestantes. Así, aunque habían sido utilizados en estudios con mujeres embarazadas, deberían haber sido validados previamente para su uso en esta población, en especial teniendo en cuenta que tienen características diferenciales de las mujeres no gestantes.

### PREGUNTA 2

Los autores comentan que el objetivo del estudio fue validar la versión española de los instrumentos de medida UDI-6 e IIQ-7, para el estudio de la incontinencia urinaria y de su repercusión en la calidad de vida de las mujeres gestantes.

Dado que la versión original de estos cuestionarios había sido validada, ¿cree que es suficiente con traducirlos al español o es necesario validar su versión española?

No podemos estar seguros de que la simple traducción al español proporcione cuestionarios con propiedades psicométricas similares a las versiones originales, ya que, además de la calidad de la traducción, la validez de una medición depende de la población concreta en que se ha estudiado, por lo que es necesario validar las versiones en diferentes idiomas, dado que se relacionan con poblaciones de características epidemiológicas y culturales diferentes. Además, debemos tener en cuenta el comentario de la pregunta anterior sobre la necesidad de la validación de los cuestionarios en población gestante.

### PREGUNTA 3

Los autores informan del proceso seguido para la elaboración de las versiones españolas de los cuestionarios: «Dos empresas profesionales de traducción de inglés a español, no familiarizadas con el uso del IIQ-7 y el UDI-6, trabajaron de forma independiente para producir dos versiones previas de cada cuestionario. En una reunión entre autores y traductores se llegó a una primera versión consensuada que se utilizó para la retrotraducción por un nativo inglés bilingüe. Esta nueva versión en inglés fue comparada con la original. Para valorar la inteligibilidad y sencillez del lenguaje, diez gestantes con síntomas de incontinencia urinaria autocumplimentaron la versión española. En una reunión final de los autores se aceptó la versión definitiva».

**¿Cómo se denomina este proceso de adaptación del cuestionario a otro idioma? ¿Cuál es su finalidad? ¿Le parece adecuado el procedimiento seguido por los investigadores?**

El método utilizado se conoce como traducción-retraducción. Es el procedimiento más adecuado para obtener una versión en otro idioma equivalente al desarrollado en el país de origen. Esta adecuación del cuestionario no puede limitarse a una simple traducción, sino que debe seguir una metodología que asegure la equivalencia conceptual y semántica con el original y la comprensión por los pacientes de la versión adaptada. Por ello se realiza la traducción-retraducción por personas bilingües, seguido de un análisis de la nueva versión para detectar discrepancias y de la comprobación con un grupo de pacientes de su comprensión y aceptabilidad, tal como hacen los investigadores. Una vez obtenida la versión final, deben comprobarse sus propiedades métricas y su validez y fiabilidad.

### PREGUNTA 4

En relación con las participantes en el estudio, los autores dicen que: «Se obtuvo una muestra de conveniencia reclutada de forma consecutiva [...]. La muestra la componen 181 gestantes que cumplieron los criterios de inclusión: estar en el tercer trimestre del embarazo, acudir a las clases de educación maternal y aceptar participar en el estudio».

**¿Cuál es la población de estudio? ¿Cree que es la adecuada?**

La población de estudio está definida por los siguientes criterios:

- Estar en el tercer trimestre del embarazo.
- Acudir a las clases de educación maternal.
- Aceptar participar en el estudio.

Lo ideal es que la población en que se valida un cuestionario sea similar a aquella en que se desea utilizar, por lo que los criterios utilizados en este estudio parecen correctos. De todas formas, debemos tener en cuenta que la finalidad del estudio de validación no es conocer la prevalencia de incontinencia urinaria, sino la validez y utilidad del cuestionario, por lo que puede desearse incluir intencionadamente sujetos de diferentes características con la finalidad de evaluar si las características del cuestionario son similares en todos ellos.

### PREGUNTA 5

**Se incluyeron 181 mujeres. Todas ellas cumplimentaron el cuestionario UDI-6.**

Los autores informan de que la tasa máxima de no respuesta fue del 3,3% (6/181) en el ítem «Indique si tiene el siguiente problema y, si es así, cuánto le molesta: pérdida de orina en pequeñas cantidades (es decir, gotas)». El cuestionario lo completaron el 92% de las mujeres (167/181), 11 no respondieron un ítem y 3 no respondieron dos ítems.

**¿Cómo interpreta estos resultados?**

El análisis de los porcentajes de no respuesta de cada ítem y el porcentaje de mujeres que no contestaron a alguno de ellos permite evaluar la facilidad de cumplimentación y la aceptabilidad de los cuestionarios, que pueden considerarse muy buenas en el caso del cuestionario UDI-6.

## PREGUNTA 6

Todas las mujeres que formaron parte del estudio, tras ser informadas sobre el objetivo de éste y dar su consentimiento, llenaron ambos cuestionarios en ese momento y transcurridas 2 semanas.

**¿Qué cree que pretenden estudiar los investigadores al repetir la administración del cuestionario? ¿Por qué se realiza pasadas 2 semanas?**

Los investigadores buscan evaluar la repetibilidad de los instrumentos, es decir, si proporcionan resultados similares cuando se administran en dos ocasiones diferentes a las mismas personas. Éste es un aspecto importante de la fiabilidad de un cuestionario (se denomina también fiabilidad test-retest). Para ser útil, un cuestionario debe tener una buena repetibilidad. Si cada vez que se administra proporciona resultados diferentes, será difícil interpretarlos.

Los autores utilizan un intervalo de 2 semanas entre ambas administraciones al considerar que es un período lo suficientemente largo para que las mujeres no recordasen las respuestas dadas en la primera administración y lo suficientemente corto como para reducir al mínimo la probabilidad de que se produjera algún cambio en cualquiera de las dimensiones estudiadas (si hubiera cambios, los resultados podrían ser diferentes y no podría evaluarse la repetibilidad, ya que las condiciones de administración serían diferentes).

## PREGUNTA 7

**¿Cree que podría evaluarse la fiabilidad interobservador?**

La fiabilidad interobservador mide el grado de concordancia entre dos observadores (entrevistadores) diferentes. Dado que se trata de cuestionarios autoadministrados, en que no existen observadores, no tiene sentido evaluar este aspecto. De hecho, ésta es precisamente una de las ventajas de los cuestionarios autoadministrados.

## PREGUNTA 8

Los autores dicen: «La fiabilidad del UDI-6 y del IIQ-7 se evaluó mediante el cálculo del coeficiente de correlación intraclass (CCI) para variables cuantitativas continuas y del índice Kappa para las variables categóricas. Para la interpretación del índice Kappa se siguió la clasificación de Fleiss ( $>0,75$  acuerdo excelente,  $\geq 0,40$  acuerdo bueno y  $<0,40$  pobre acuerdo). Para la interpretación del CCI se utilizó la clasificación de Landis ( $>0,91$  concordancia muy buena,  $\geq 0,71$  buena,  $\geq 0,51$  moderada,  $\geq 0,31$  mediocre y  $<0,30$  mala o muy mala)».

**¿Cree que este análisis es correcto?**

Sí. El análisis de la fiabilidad se basa en la evaluación de la concordancia entre las respuestas dadas en las dos administraciones de los cuestionarios. Esta concordancia se mide mediante pruebas estadísticas diferentes: índice Kappa cuando se trata de variables categóricas, y coeficiente de correlación intraclass cuando son cuantitativas.

## PREGUNTA 9

En los resultados, los autores informan de que los índices Kappa fueron superiores a 0,75 en todos los diagnósticos, y que el CCI se situó por encima de 0,71 en el UDI-6 y de 0,91 en el IIQ-7:

		Kappa	p
UDI-6	Presencia de incontinencia urinaria	0,966	<0,01
	Tipos de incontinencia urinaria:		
	Incontinencia de esfuerzo	0,903	<0,01
	Incontinencia de urgencia	0,938	<0,01
	Incontinencia mixta	0,856	<0,01

## Métodos de investigación clínica y epidemiológica

		CCI	p
UDI-6	Gravedad de los síntomas	0,902	<0,01
	Subescala síntomas irritativos	0,899	<0,01
	Subescala síntomas de estrés	0,898	<0,01
	Subescala obstrucción/dolor	0,812	<0,01
IIQ-7	Calidad de vida	0,954	<0,001

### ¿Cómo interpreta estos resultados?

Ambos cuestionarios muestran una buena repetibilidad, tal como indican los elevados grados de concordancia observados entre ambas administraciones de los cuestionarios.

### PREGUNTA 10

Otro aspecto de la fiabilidad de un cuestionario es la consistencia interna. ¿Cree que puede analizarse en este estudio? Si es posible, ¿qué prueba utilizaría?

La consistencia interna se refiere al grado en que las respuestas a los ítems que miden un mismo atributo presentan homogeneidad entre ellas. La evaluación de la consistencia interna sólo requiere la administración del cuestionario en una única ocasión, por lo que puede ser realizada en este estudio. La técnica estadística para su análisis es el alfa de Cronbach.

En relación con ello, los autores dicen:

«La consistencia interna, que evalúa la correlación general entre los elementos dentro de una escala, se calculó con el coeficiente alfa de Cronbach; valores de 0,70 o superiores se aceptan como indicador de buena consistencia interna».

### PREGUNTA 11

Los valores del coeficiente alfa de Cronbach observados son los siguientes:

		Cronbach
UDI-6	Gravedad de los síntomas	0,667
	Subescala síntomas irritativos	0,591
	Subescala síntomas de estrés	0,799
	Subescala obstrucción/dolor	0,395
IIQ-7	Calidad de vida	0,910

### ¿Cómo interpreta estos resultados?

En relación con estos resultados, los autores comentan: «El coeficiente alfa de Cronbach de 0,667 para el UDI-6 y de 0,910 para el IIQ-7 fundamenta una buena consistencia interna de ambos instrumentos, con valores próximos a los hallados en el estudio de validación al turco y al chino, pero superiores a los obtenidos en el estudio de validación al árabe. Las subescalas de síntomas irritativos y de obstrucción/dolor no alcanzan valores de buena consistencia interna, 0,591 y 0,395, respectivamente. Este resultado podría estar relacionado con el número de ítems que las componen y con el hecho de que las pérdidas con urgencia y los síntomas de obstrucción/dolor tienen menos prevalencia en la gestación que los síntomas de incontinencia urinaria de esfuerzo».

### PREGUNTA 12

En la descripción del método, los autores dicen: «Junto con los cuestionarios UDI-6 e IIQ-7, las mujeres autocumplimentaron la versión española del Consultation on Incontinence Questionnaire Urinary Incontinence Short Form (ICIQ-SF), que mide síntomas y calidad de vida, y que ha demostrado unas buenas propiedades psicométricas».

### ¿Cuál es la utilidad de la administración de este cuestionario?

Los autores utilizan este cuestionario como estándar de referencia de la presencia y del tipo de incontinencia urinaria, de manera que puedan comparar los resultados de los cuestionarios que están validando con este estándar. Se trataría, por tanto, de la validez de criterio de los cuestionarios.

### PREGUNTA 13

#### ¿Cree que el cuestionario ICIQ-SF es un buen estándar de referencia para evaluar la validez de los otros cuestionarios?

El estándar de referencia es muy importante, al igual que ocurre en los estudios que evalúan una prueba diagnóstica. Dado que ha de permitir conocer quién tiene y quién no tiene la enfermedad o problema de salud de interés, ha de ser un buen estándar ampliamente reconocido como tal.

En general, los mejores estándares de referencia suelen ser medidas «objetivas», habitualmente variables de laboratorio o resultados de exploraciones complementarias, aunque cuando ninguna de ellas se considera suficientemente buena, puede recurrirse a otro tipo de medidas.

En relación con este aspecto, los autores comentan: «La principal dificultad que se presenta en la validación de escalas de síntomas y calidad de vida estriba en la difícil comparación con un patrón de referencia que facilite el análisis de validez. Aunque no hay consenso sobre el mejor criterio externo, la International Continence Society (ICS) recomienda utilizar como criterio externo para la clasificación clínica de la incontinencia urinaria el estudio urodinámico, aunque se acepta que no es un buen criterio para valorar la calidad de vida. Éste es el criterio utilizado en los estudios de validación en árabe y turco. Otros criterios utilizados son el *pad-test*, el número de episodios de incontinencia u otro cuestionario. La condición de gestantes de las mujeres del estudio dificulta aún más la adopción del criterio externo; que sepamos, ninguno de los criterios diagnósticos recomendados para el estudio de la incontinencia urinaria, urodinamia, *pad-test* u otros cuestionarios han sido validados en mujeres gestantes.

» Se ha recurrido a su validación respecto al ICIQ-SF porque, aunque no se ha validado en gestantes, sí ha sido validada su versión española, se ha utilizado en estudios epidemiológicos de incontinencia urinaria en gestantes y es un instrumento de medida aceptable en embarazadas».

### PREGUNTA 14

#### De las 181 mujeres embarazadas incluidas en el estudio, 92 (50,8%) presentan incontinencia urinaria según el cuestionario ICIQ-SF. ¿A qué tipo de medida corresponde este valor?

Dado que se trata de un estudio transversal realizado sobre una muestra de 181 mujeres embarazadas, de las que 92 presentan el problema de salud de estudio, el valor 50,8% corresponde a la prevalencia de la incontinencia urinaria entre las mujeres estudiadas.

### PREGUNTA 15

#### Los autores informan de que el valor del UDI-6 difirió significativamente entre mujeres con y sin incontinencia urinaria, según el cuestionario ICIQ-SF ( $p < 0,01$ ). ¿Cómo interpreta este dato?

Si asumimos que el cuestionario ICIQ-SF clasifica correctamente a las mujeres embarazadas con y sin incontinencia urinaria, puede decirse que los valores del cuestionario UDI-6 discriminan adecuadamente ambos grupos de mujeres.

### PREGUNTA 16

#### La sensibilidad del UDI-6 para la presencia de incontinencia urinaria en relación con el ICIQ-SF fue del 98,9%. ¿Qué significa este resultado?

Significa que, de las mujeres embarazadas con incontinencia urinaria (según el ICIQ-SF), prácticamente el 99% obtendría un resultado positivo con el cuestionario UDI-6.

### PREGUNTA 17

#### La especificidad del UDI-6 para la presencia de incontinencia urinaria en relación con el ICIQ-SF fue del 86,5%. ¿Qué significa este resultado?

## Métodos de investigación clínica y epidemiológica

Significa que, de las mujeres embarazadas sin incontinencia urinaria (según el ICIQ-SF), el 86,5% obtendría un resultado negativo con el cuestionario UDI-6.

### PREGUNTA 18

**El valor predictivo positivo del UDI-6 para la presencia de incontinencia urinaria en relación con el ICIQ-SF fue del 88,3%. ¿Qué significa este resultado?**

Significa que, de las mujeres embarazadas que presentan un resultado positivo en el cuestionario UDI-6, el 88,3% tiene incontinencia urinaria (según ICIQ-SF).

### PREGUNTA 19

**¿Cree que el cuestionario UDI-6 tiene validez de criterio?**

La validez de criterio del cuestionario UDI-6 se ha evaluado en relación con el cuestionario ICIQ-SF. En relación con la presencia de incontinencia urinaria determinada según este criterio, los valores de sensibilidad, especificidad y valor predictivo indican que el cuestionario UDI-6 tiene una buena validez de criterio.

### PREGUNTA 20

**¿Cuál es su conclusión del estudio?**

La conclusión de los autores es la siguiente: «Las versiones españolas del IIQ-7 y del UDI-6 son fácilmente comprendidas por las gestantes, lo que facilita su autocomplimentación. Los resultados del estudio demuestran que los dos cuestionarios constituyen un instrumento fiable, consistente y válido para evaluar los síntomas urogenitales y su impacto en la calidad de vida de las mujeres embarazadas».

## PREGUNTAS ABIERTAS DE CÁLCULO

### PREGUNTA 1

Los resultados de un estudio de incidencia de una enfermedad se presentan en la siguiente tabla. Todas las personas se siguen durante 2 años.

Exposición	Número de personas	Número de casos
Positiva	500	10
Negativa	10.000	10

¿Cuál es la incidencia por 1.000 personas-año en aquellas personas expuestas?

Respuesta: Según la tabla, de las 500 personas expuestas, 10 desarrollan la enfermedad. Como las 500 personas son seguidas durante 2 años, el número de personas-año expuestas es 1.000. Por tanto, la incidencia entre personas expuestas es de 10 por 1.000 personas-año.

### PREGUNTA 2

Siguiendo con el mismo estudio de la anterior pregunta, ¿cuál es la incidencia entre las personas no expuestas?

Respuesta: Siguiendo el mismo razonamiento de la pregunta anterior, el número de personas-año expuestas es de  $10.000 \times 2 = 20.000$ . Por tanto, la incidencia es de  $\frac{10 \text{ nuevos casos}}{20.000} = 0,5 \text{ casos por 1.000 personas-año}$ .

### PREGUNTA 3

Siguiendo con la tabla de la pregunta 1, ¿cuál es el riesgo relativo entre personas expuestas y no expuestas?

Respuesta: Incidencia entre expuestas/incidencia entre no expuestas;  $10/0,5 = 20$ .

### PREGUNTA 4

Durante la investigación de una intoxicación alimentaria en un banquete, se obtiene la historia de ingesta de un determinado alimento de todos los asistentes. ¿Cuál es el alimento que probablemente causó la intoxicación?

Alimento	Afectados que comieron el alimento	Afectados que no comieron el alimento
A	61	48
B	70	35
C	40	50
D	73	10

Respuesta: El alimento D. La identificación del alimento se realiza comparando la proporción entre los que comen y no comen el alimento específico. Aunque en ocasiones esta medida se denomina tasa de ataque, en realidad es una proporción.

### PREGUNTA 5

En un estudio que evalúa una nueva prueba diagnóstica se realiza el test en 100 personas con la enfermedad y en 100 personas sin la enfermedad. El resultado se muestra en la siguiente tabla:

Prueba	Con enfermedad	Sin enfermedad
Positiva	91	12
Negativa	9	88

Calcular la sensibilidad del test.

Respuesta:  $91/100 = 0,91$

## Métodos de investigación clínica y epidemiológica

### PREGUNTA 6

En un estudio que evalúa una nueva prueba diagnóstica se realiza el test en 100 personas con la enfermedad y en 100 personas sin la enfermedad. El resultado se muestra en la siguiente tabla:

Prueba	Con enfermedad	Sin enfermedad
Positiva	91	12
Negativa	9	88

Calcular la especificidad del test.

Respuesta:  $88/100 = 0,88$

### PREGUNTA 7

En un estudio que evalúa una nueva prueba diagnóstica se realiza el test en 100 personas con la enfermedad y en 100 personas sin la enfermedad. El resultado se muestra en la siguiente tabla:

Prueba	Con enfermedad	Sin enfermedad
Positiva	91	12
Negativa	9	88

Calcular el valor predictivo positivo.

Respuesta:  $91/(91+12)$

### PREGUNTA 8

En un estudio que evalúa una nueva prueba diagnóstica se realiza el test en 100 personas con la enfermedad y en 100 personas sin la enfermedad. El resultado se muestra en la siguiente tabla:

Prueba	Con enfermedad	Sin enfermedad
Positiva	91	12
Negativa	9	88

Calcular el valor predictivo negativo.

Respuesta:  $88/(88+9)$

### PREGUNTA 9

La incidencia de cáncer de pulmón en la comunidad es de 120/100.000 personas-año en los fumadores y de 10/100.000 personas-año en los no fumadores. La prevalencia de consumo de tabaco es del 20% en la comunidad.

Calcular el riesgo de enfermedad en los fumadores comparado con los no fumadores.

Respuesta:  $\frac{120/100.000}{10/100.000} = 12$

### PREGUNTA 10

La incidencia de cáncer de pulmón en la comunidad es de 120/100.000 personas-año en los fumadores y de 10/100.000 personas-año en los no fumadores. La prevalencia de consumo de tabaco es del 20% en la comunidad.

Calcular el porcentaje de casos de cáncer de pulmón que pueden atribuirse al consumo de tabaco.

Respuesta: El riesgo atribuible o diferencia de incidencia sería:

$(120/100.000) - (10/100.000) = 110/100.000$

El porcentaje de riesgo atribuible sería:

$$[(\text{Incidencia en expuestos} - \text{incidencia no expuestos})/\text{incidencia expuestos}] \times 100 = 92\%$$

### PREGUNTA 11

La incidencia de cáncer de pulmón en la comunidad es de 120/100.000 personas-año en los fumadores y de 10/100.000 personas-año en los no fumadores. La prevalencia de consumo de tabaco es del 20% en la comunidad.

Si la prevalencia de consumo de tabaco descendiera hasta el 10%, ¿cuál sería el exceso de incidencia que se evitaría en la comunidad?

**Respuesta:** Si la prevalencia se redujera un 10%, se evitarían 11/100.000 casos de cáncer de pulmón debido al consumo de tabaco.

$$(\text{Riesgo atribuible} \times \text{prevalencia de exposición inicial [20\%]}) - (\text{Riesgo atribuible} \times \text{reducción de prevalencia [10\%]})$$

### PREGUNTA 12

En una muestra de 1.000 personas, 100 tienen la enfermedad en estudio. Después de un año de seguimiento, 90 desarrollan la enfermedad.

Calcular la prevalencia de la enfermedad al inicio del estudio.

$$\text{Respuesta: } 100/1.000 = 0,10$$

### PREGUNTA 13

En una muestra de 1.000 personas, 100 tienen la enfermedad en estudio. Después de un año de seguimiento, 90 desarrollan la enfermedad.

Calcular la incidencia de la enfermedad.

$$\text{Respuesta: } 90/1.000 - 100 = 0,10$$

En el denominador hemos de restar las personas que ya tenían la enfermedad al inicio del estudio y, por tanto, no entran en el seguimiento.

### PREGUNTA 14

En el estudio de una enfermedad crónica, se elige una muestra de 1.000 personas, de las cuales 100 tienen la enfermedad. Al finalizar el estudio, 100 personas más han desarrollado la enfermedad.

Calcular la prevalencia de la enfermedad.

$$\text{Respuesta: } (100 + 100)/1.000 = 0,20$$

### PREGUNTA 15

En un estudio se elige una muestra de 100 personas con la enfermedad y 100 personas sin la enfermedad. Entre los enfermos, 60 están expuestos al factor de estudio, mientras que sólo 20 personas no enfermas están expuestas.

Prueba	Casos	Controles
Expuestas	60	20
No expuestas	40	80
	100	100

Calcular la odds ratio.

$$\text{Respuesta: Odds ratio} = (60 \times 80) / (40 \times 20) = 6$$

## Métodos de investigación clínica y epidemiológica

### PREGUNTA 16

En un ensayo clínico se asignan aleatoriamente 2.000 pacientes: 1.000 que reciben un nuevo fármaco y 1.000 que reciben un placebo. Al cabo de un año de seguimiento, 200 pacientes con el nuevo fármaco y 100 que recibieron el placebo se han curado.

Calcular la incidencia acumulada de curación en los que recibieron el tratamiento.

Respuesta:  $200/1.000 = 0,20$

### PREGUNTA 17

En un ensayo clínico se asignan aleatoriamente 2.000 pacientes: 1.000 que reciben un nuevo fármaco y 1.000 que reciben un placebo. Al cabo de un año de seguimiento, 200 pacientes con el nuevo fármaco y 100 que recibieron el placebo se han curado.

Calcular la incidencia acumulada de curación en el grupo placebo.

Respuesta:  $100/1.000 = 0,10$

### PREGUNTA 18

En un ensayo clínico se asignan aleatoriamente 2.000 pacientes: 1.000 que reciben un nuevo fármaco y 1.000 que reciben un placebo. Al cabo de un año de seguimiento, 200 pacientes con el nuevo fármaco y 100 que recibieron el placebo se han curado.

Calcular la diferencia de incidencia entre el grupo que recibe el nuevo fármaco y el grupo placebo.

Respuesta:  $Diferencia\ de\ incidencia = 0,20 - 0,10 = 0,10$

### PREGUNTA 19

En un ensayo clínico se asignan aleatoriamente 2.000 pacientes: 1.000 que reciben un nuevo fármaco y 1.000 que reciben un placebo. Al cabo de un año de seguimiento, 200 pacientes con el nuevo fármaco y 100 que recibieron el placebo se han curado.

Calcular en número de personas que es necesario tratar (NNT).

Respuesta:  $NNT = 1/diferencia\ de\ incidencia = 1/0,10 = 10$

### PREGUNTA 20

En un ensayo clínico se asignan aleatoriamente 2.000 pacientes: 1.000 que reciben un nuevo fármaco y 1.000 que reciben un placebo. Al cabo de un año de seguimiento, 200 pacientes con el nuevo fármaco y 100 que recibieron el placebo se han curado.

Calcular la probabilidad de curación con el nuevo fármaco en relación con el placebo.

Respuesta: Se calcula dividiendo la incidencia de curación con el nuevo fármaco (0,20) por la incidencia de curación con el placebo (0,10).

$0,20/0,10 = 2$

### PREGUNTA 21

En un ensayo clínico se asignan aleatoriamente 2.000 pacientes: 1.000 que reciben un nuevo fármaco y 1.000 que reciben un placebo. Al cabo de un año de seguimiento, 200 pacientes con el nuevo fármaco y 100 que recibieron el placebo se han curado.

Calcular la reducción de riesgo relativo (RRR).

Respuesta:  $RRR = (incidencia\ con\ el\ nuevo\ tratamiento - incidencia\ con\ placebo) / incidencia\ nuevo\ tratamiento$ .

$0,10/0,20 = 0,50$

### PREGUNTA 22

Un estudio pretende estimar el porcentaje de pacientes fumadores en un centro de salud. A partir de estudios previos, se estima que debe estar alrededor del 40% ( $p=0,40$ ). Se quiere realizar la estimación con una previsión del 4% ( $i=0,04$ ) y una confianza del 95%.

Calcular el número de sujetos necesarios.

**Respuesta:** 576

### PREGUNTA 23

Siguiendo con el estudio de la pregunta 22, se desea realizar la estimación con una previsión del 1% ( $i=0,01$ ).

Calcular el número de sujetos necesarios.

**Respuesta:** 9.220

### PREGUNTA 24

Siguiendo con el ejemplo de la pregunta 22, se desea realizar la estimación con una previsión del 10% ( $i=0,10$ ).

Calcular el número de sujetos necesarios.

**Respuesta:** 92

### PREGUNTA 25

Se desea estudiar la presión arterial diastólica en un grupo de pacientes del que se conoce que la desviación estándar está alrededor de 25 mmHg. Se quiere realizar la estimación con una confianza del 95% y una previsión de 5 mmHg ( $i=5$ ).

Calcular el número de sujetos necesarios.

**Respuesta:** A partir de la tabla B de los anexos, el número de sujetos es 96.

### PREGUNTA 26

En el tratamiento estándar de una enfermedad se consigue un 40% de curaciones. Con un nuevo tratamiento se espera que, como mínimo, se consiga un 50% de curaciones. El riesgo alfa que se acepta es del 5%, con una potencia estadística del 80%.

Calcular el número de sujetos necesarios.

**Respuesta:** A partir de la tabla C, 387 sujetos.

### PREGUNTA 27

Siguiendo con el ejemplo de la pregunta 26, se desea estudiar el número de sujetos con una potencia del 90%.

Calcular el tamaño de la muestra.

**Respuesta:** 518

### PREGUNTA 28

Siguiendo con el ejemplo de la pregunta 26, se desea detectar una diferencia mínima con el nuevo tratamiento del 20%, con una potencia del 90% y un riesgo alfa del 5%.

Calcular el número de sujetos necesarios.

**Respuesta:** 97

### PREGUNTA 29

Los resultados de un estudio que evalúa una nueva prueba diagnóstica se presentan en la siguiente tabla:

Prueba	Enfermedad	No enfermedad
Positiva	80	10
Negativa	20	90
	100	100

## Métodos de investigación clínica y epidemiológica

Calcular la razón de probabilidad positiva.

**Respuesta:** La razón de probabilidad positiva es  $(\text{sensibilidad})/(1 - \text{especificidad})$ :  $0,80/(1 - 0,90) = 8$

### PREGUNTA 30

Los resultados de un estudio que evalúa una nueva prueba diagnóstica se presentan en la siguiente tabla:

Prueba	Enfermedad	No enfermedad
Positiva	80	10
Negativa	20	90
	100	100

Calcular la razón de probabilidad negativa.

**Respuesta:** La razón de probabilidad negativa es  $(1 - \text{sensibilidad})/\text{especificidad}$ :  $0,20/0,90$

## PREGUNTAS DE MÚLTIPLE ELECCIÓN

### PREGUNTA 1

La *odds ratio* estima:

- A) La fuerza de la asociación entre una exposición y una enfermedad.
- B) La proporción de casos atribuible a la exposición.
- C) La diferencia de riesgos entre expuestos y no expuestos a un factor.
- D) La magnitud de un problema de salud en una comunidad.
- E) El número de casos que se podrían prevenir si se eliminara la exposición.

*Respuesta: A. La odds ratio mide la fuerza de la asociación entre la exposición y la enfermedad, pero no permite estimar la magnitud de un problema de salud en una comunidad. El resto de las opciones hacen referencia al impacto que tendría sobre la población el control o la eliminación de una exposición. La fuerza de la asociación entre el factor y la enfermedad puede ser baja, pero si la exposición al factor de riesgo en la población es muy frecuente, su repercusión en el ámbito de la salud pública puede ser importante.*

### PREGUNTA 2

¿En cuál de los siguientes estudios no se puede asegurar una adecuada secuencia temporal entre una enfermedad y su posible causa?

- A) Ensayo clínico.
- B) Cohortes prospectivas.
- C) Cohortes retrospectivas.
- D) Casos y controles dentro de una cohorte.
- E) Estudio de casos y controles con casos prevalentes.

*Respuesta: E. La característica principal de los estudios de casos y controles con casos prevalentes es que la presencia de la exposición (causa o factor de riesgo) y de la enfermedad se observa simultáneamente, lo que condiciona una falta de secuencia temporal entre la medición de un factor de riesgo y la aparición de una enfermedad. Esta ambigüedad temporal dificulta en muchas ocasiones la interpretación de una posible relación causa-efecto. En el resto de los estudios, al existir un seguimiento de los pacientes, no se presenta esta falta de secuencia temporal.*

### PREGUNTA 3

¿Qué tipo de estudio es el más adecuado para estimar la prevalencia de una enfermedad en una población?

- A) Casos y controles.
- B) Transversal.
- C) Cohorte prospectiva.
- D) Cohorte retrospectiva.
- E) Ensayo clínico.

*Respuesta: B. El objetivo principal de los estudios transversales es estudiar la prevalencia (proporción de individuos) que presentan una enfermedad en una población.*

### PREGUNTA 4

Si de los 1.000 pacientes hipertensos que residen en una zona básica de salud, 60 desarrollan cardiopatía isquémica después de un año de seguimiento, ¿cuál de las siguientes afirmaciones es correcta?

- A) La prevalencia es del 60 por 1.000.
- B) La incidencia acumulada es del 60 por 1.000.
- C) El riesgo relativo es del 60 por 1.000.
- D) El riesgo atribuible es del 60 por 1.000.
- E) La fracción etiológica es del 60 por 1.000.

*Respuesta: B. La incidencia se define como el número de casos nuevos de una enfermedad que se desarrollan en una población de riesgo durante un período de seguimiento. La prevalencia consiste en todos los casos que existen en la población, no sólo los casos nuevos. El resto son medidas relativas que relacionan varias medidas de incidencia.*

### PREGUNTA 5

¿Cuál de las siguientes medidas estima el riesgo de desarrollar una enfermedad?

- A) La incidencia acumulada.
- B) La diferencia de incidencias.
- C) El riesgo relativo.
- D) La fracción atribuible.
- E) La prevalencia.

*Respuesta: A. La incidencia acumulada mide la probabilidad (o riesgo) de que un individuo desarrolle una enfermedad en un tiempo determinado.*

### PREGUNTA 6

Si extrae una muestra aleatoria de individuos de una comunidad y estima cuántos presentan una determinada enfermedad en un momento concreto del tiempo, usted está efectuando un estudio:

- A) De cohortes prospectivas.
- B) De cohortes retrospectivas.
- C) Transversal.
- D) Casos y controles.
- E) Ecológico.

*Respuesta: C. La descripción del proceso que se presenta en el enunciado corresponde a los estudios transversales, cuyo principal objetivo es estimar la prevalencia de la enfermedad en un momento concreto del tiempo. Para estudiar la incidencia o el riesgo de enfermar debe existir seguimiento de los sujetos.*

### PREGUNTA 7

¿Cuál de los siguientes es el estudio de elección para analizar la efectividad de un nuevo tratamiento?

- A) Transversal.
- B) Casos y controles.
- C) Cohortes.
- D) Ensayo clínico aleatorio.
- E) Serie de casos clínicos.

*Respuesta: D. La evaluación de la eficacia y la efectividad de una intervención, ya sea farmacológica o una prueba diagnóstica, debe hacerse preferentemente mediante estudios experimentales. De los estudios mencionados, el único experimental es el ensayo clínico aleatorio.*

### PREGUNTA 8

El valor predictivo positivo de una prueba diagnóstica es:

- A) La probabilidad de que la prueba dé un resultado positivo cuando se aplica a un enfermo.
- B) La probabilidad de que la prueba dé un resultado negativo cuando se aplica a un no enfermo.
- C) La probabilidad de que un individuo con un resultado positivo tenga la enfermedad.
- D) La probabilidad de que un individuo con un resultado negativo no tenga la enfermedad.
- E) La probabilidad de que la prueba dé un resultado positivo cuando se aplica a un no enfermo.

*Respuesta: C. El valor predictivo positivo se define como la probabilidad de que un individuo con un resultado positivo de la prueba diagnóstica tenga la enfermedad.*

### PREGUNTA 9

Una prueba diagnóstica que tiene una sensibilidad del 80%:

- A) Detectará correctamente a un 80% de los individuos sin la enfermedad.
- B) No detectará correctamente a un 20% de los individuos sin la enfermedad.
- C) No detectará a un 20% de los individuos con la enfermedad.
- D) La exactitud de la prueba es del 80%.
- E) El 80% de los individuos que tienen la prueba positiva tienen la enfermedad.

**Respuesta:** C. La sensibilidad es la probabilidad de que una prueba diagnóstica detecte correctamente a los individuos con la enfermedad. En la pregunta, significa que la prueba detectará correctamente al 80% de los individuos con la enfermedad, o lo que es lo mismo, no detectará al 20% de los individuos que tienen la enfermedad.

### PREGUNTA 10

En una prueba diagnóstica cuya especificidad es del 90%, ¿cuál de las siguientes afirmaciones es cierta?

- A) La probabilidad de un resultado falso positivo es del 10%.
- B) La probabilidad de un resultado falso negativo es del 10%.
- C) La sensibilidad de la prueba es del 10%.
- D) El valor predictivo positivo de la prueba es del 10%.
- E) El valor predictivo negativo de la prueba es del 10%.

**Respuesta:** A. La especificidad de una prueba diagnóstica es la probabilidad de que su resultado sea negativo cuando realmente no existe la enfermedad. El complementario de la especificidad es el tanto por ciento de falsos positivos de la prueba diagnóstica. Así pues, si la especificidad de una prueba es del 90%, el tanto por ciento de falsos positivos es del 10%.

### PREGUNTA 11

¿Cuál de las siguientes medidas epidemiológicas utilizaría para cuantificar el impacto potencial de un programa preventivo en la población?

- A) Riesgo relativo.
- B) Odds ratio.
- C) Razón de prevalencias.
- D) Disminución de la prevalencia.
- E) Riesgo atribuible.

**Respuesta:** E. El riesgo atribuible o diferencia de incidencias es una medida de impacto que muestra el exceso de casos que están asociados a una exposición y que podrían evitarse si ésta se eliminara. El riesgo atribuible depende no sólo del riesgo relativo o la odds ratio, sino también de la frecuencia de la enfermedad. La disminución de la prevalencia no depende sólo de cambios en la incidencia de la enfermedad, sino también de su duración.

### PREGUNTA 12

Si desea evaluar la evidencia científica de que la disminución de las cifras de presión arterial en pacientes hipertensos causa una disminución en la incidencia de la enfermedad cardiovascular, usted buscará estudios con el siguiente diseño:

- A) Casos y controles.
- B) Ensayo clínico.
- C) Cohortes.
- D) Transversal.
- E) Serie de casos clínicos.

**Respuesta:** B. El objetivo principal de los estudios experimentales es evaluar la eficacia de cualquier intervención preventiva, curativa o rehabilitadora. En el ensayo clínico, al existir la asignación aleatoria del tratamiento, se tiene un alto grado de control de la situación, que proporciona, en el supuesto de que exista una asociación entre el tratamiento y el resultado observado, una fuerte probabilidad de que dicha relación sea causal. El resto de los estudios son observacionales, en los que no es posible la asignación aleatoria.

### PREGUNTA 13

¿A cuál de los siguientes conceptos se refiere la expresión «análisis por intención de tratar» en un ensayo clínico aleatorio?

- A) A analizar solamente los sujetos asignados al grupo que ha recibido el tratamiento de interés.
- B) A analizar los sujetos según la intervención que han recibido en realidad.

## Métodos de investigación clínica y epidemiológica

- C) A analizar los sujetos en el grupo al que han sido asignados, independientemente del tratamiento que han recibido.
- D) A analizar solamente a los sujetos que han cumplido con la intervención de interés.
- E) A excluir del análisis las pérdidas y los abandonos de seguimiento, y los fallos de cumplimiento.

*Respuesta: C. La expresión «análisis por intención de tratar» se aplica a los ensayos clínicos en que los análisis de los datos se realizan según el tratamiento asignado inicialmente por distribución aleatoria. El análisis por intención de tratar mide la efectividad derivada de ofrecer un tratamiento a un paciente, independientemente de si después este tratamiento es tomado, tiene que ser interrumpido, etc.*

### PREGUNTA 14

En un estudio que evalúa una prueba diagnóstica, se estima que su valor predictivo positivo es del 0,80. ¿Cómo interpreta este resultado?

- A) La probabilidad que la prueba dé un resultado positivo cuando se aplica a un enfermo es del 0,80.
- B) La probabilidad de que la prueba dé un resultado negativo cuando se aplica a un no enfermo es del 0,80.
- C) La probabilidad de que un individuo con un resultado positivo tenga la enfermedad es del 0,80.
- D) La probabilidad de que un individuo con un resultado negativo no tenga la enfermedad es del 0,80.
- E) La probabilidad de que la prueba dé un resultado positivo cuando se aplica a un no enfermo es del 0,80.

*Respuesta: C. El valor predictivo positivo se define como la probabilidad de que un individuo con un resultado positivo de la prueba diagnóstica tenga la enfermedad.*

### PREGUNTA 15

Si una prueba diagnóstica que tiene una sensibilidad del 90% y una especificidad también del 90% se aplica a una población de 200 individuos con una prevalencia de enfermedad del 50%. ¿Cuál será el valor predictivo positivo?

- A) 90%.
- B) 80%.
- C) 70%.
- D) 60%.
- E) 50%.

*Respuesta: A. El número total de resultados positivos en esta muestra de individuos será de 100 (90 verdaderos positivos más 10 falsos positivos). De estos 100 individuos con resultado positivo, 90 tienen realmente la enfermedad, por lo que el valor predictivo positivo será del 90%.*

### PREGUNTA 16

¿Cuál de las siguientes afirmaciones sobre la incidencia acumulada es cierta?

- A) Estima el riesgo de desarrollar una enfermedad.
- B) Es una tasa.
- C) Se interpreta como el número de personas que tienen la enfermedad en un momento del tiempo.
- D) Para su estimación no es necesario que haya seguimiento de los pacientes.
- E) El denominador es toda la población.

*Respuesta: B. La incidencia es un concepto que se refiere al número de nuevos casos de una enfermedad que aparecen en una comunidad durante un período de tiempo determinado. Se entiende por incidencia acumulada la proporción de individuos sanos que desarrollan la enfermedad durante dicho período, por lo que, en sentido estricto, no es una tasa sino una proporción, cuyo denominador es la población en riesgo de desarrollar la enfermedad y que estima el riesgo o la probabilidad de que un individuo sano la desarrolle. Para su interpretación es necesario conocer el tiempo de observación, que se asume que es el mismo para todos los sujetos.*

**PREGUNTA 17**

¿Cuál de las siguientes afirmaciones sobre el concepto de persona-tiempo en epidemiología es falsa?

- A) Su utilización asume que el riesgo de enfermedad es constante a lo largo de todo el período de observación.
- B) Su utilización asume que la incidencia entre los que se pierden durante el seguimiento es la misma que en los que continúan.
- C) Tiene en cuenta el tiempo de seguimiento de cada sujeto.
- D) La medida persona-tiempo debe ser adecuada a la historia natural de la enfermedad.
- E) Se utiliza como denominador para el cálculo de la incidencia acumulada.

*Respuesta: E. Uno de los problemas que aparecen habitualmente al estimar la incidencia de una enfermedad es que el tiempo de observación no es el mismo para todos los sujetos estudiados. En esta situación se recurre al cálculo de la denominada densidad de incidencia, cuyo denominador se expresa en unidades de persona-tiempo, a diferencia de lo que ocurre al calcular la incidencia acumulada. Esta medida tiene en cuenta el tiempo de observación de cada sujeto, y asume que el riesgo de desarrollar la enfermedad es constante a lo largo de todo el período y que la incidencia de la enfermedad es similar entre los sujetos que se pierden durante el seguimiento y los que continúan en él. La elección de la medida de persona-tiempo debe ser adecuada a la historia natural de la enfermedad.*

**PREGUNTA 18**

¿En qué tipo de estudio la unidad de análisis son agregaciones de sujetos?

- A) Ensayos clínicos pragmáticos.
- B) Cohortes retrospectivas.
- C) Casos y controles.
- D) Estudios ecológicos.
- E) Ensayos clínicos explicativos.

*Respuesta: D. Los estudios ecológicos se caracterizan porque la unidad de análisis no es el individuo sino una agregación de ellos, habitualmente basada en criterios geográficos. Por consiguiente, el hallazgo de una asociación a este nivel no garantiza que exista también a nivel individual. La obtención de conclusiones inapropiadas en el plano individual a partir de este tipo de estudios es un error conocido como falacia ecológica. Se trata de estudios descriptivos, rápidos, económicos y fáciles de realizar.*

**PREGUNTA 19**

¿Cuál de las siguientes afirmaciones sobre las medidas epidemiológicas de asociación e impacto es falsa?

- A) El riesgo relativo es útil en investigación etiológica.
- B) El riesgo relativo depende de la frecuencia de la enfermedad.
- C) La diferencia de incidencias es útil en salud pública.
- D) El riesgo relativo indica la magnitud de la asociación entre la enfermedad y la exposición al factor de riesgo.
- E) La diferencia de incidencias depende de la frecuencia de la enfermedad.

*Respuesta: B. Las medidas epidemiológicas de asociación, como el riesgo relativo, estiman la magnitud de la asociación entre la enfermedad y la exposición al factor de riesgo, por lo que son útiles en investigación etiológica. Sin embargo, se trata de medidas adimensionales que no tienen en cuenta la frecuencia de la enfermedad, ya que sólo indican el número de veces que se incrementa el riesgo. En cambio, las medidas de impacto, como la diferencia de incidencias, estiman el exceso de riesgo asociado a la exposición, que depende de la frecuencia de la enfermedad, por lo que son medidas útiles desde el punto de vista de salud pública.*

**PREGUNTA 20**

¿Cuál de las siguientes no es una ventaja de los estudios transversales?

- A) Permiten estimar la prevalencia de una enfermedad.
- B) Permiten estudiar varias enfermedades simultáneamente.

## Métodos de investigación clínica y epidemiológica

- C) Permiten estudiar varios factores de riesgo simultáneamente.
- D) Se realizan en un corto período de tiempo.
- E) Son útiles para estudiar enfermedades poco frecuentes.

*Respuesta: E. Los estudios transversales examinan la relación entre una enfermedad y una serie de variables en una población determinada y en un momento del tiempo. Al analizar una muestra representativa de la población, permiten estimar la prevalencia de la enfermedad, así como estudiar varias enfermedades y varios factores de riesgo simultáneamente. Tienen la ventaja de que pueden realizarse en un corto período de tiempo, pero no son útiles para estudiar enfermedades poco frecuentes, ya que requerirían una muestra con un número de individuos muy grande.*

### PREGUNTA 21

¿Cuál de las siguientes no es una característica de los estudios de cohortes?

- A) Son útiles en enfermedades poco frecuentes.
- B) Son útiles en enfermedades con cortos períodos de latencia.
- C) Permiten estimar la incidencia de una enfermedad.
- D) Permiten estudiar varias enfermedades simultáneamente.
- E) Tienen una correcta secuencia temporal causa-efecto.

*Respuesta: A. Los estudios de cohortes son estudios observacionales en los que existe un seguimiento de una cohorte de sujetos expuestos a un factor de riesgo y de otra de sujetos no expuestos, para determinar la incidencia en cada uno de los grupos. Tienen la ventaja de que aseguran la correcta secuencia temporal causa-efecto y de que permiten evaluar diferentes efectos (enfermedades) de la exposición. Sin embargo, son poco útiles en enfermedades poco frecuentes y con largos períodos de latencia, ya que debería estudiarse un número grande de sujetos y seguirlos durante períodos de tiempo prolongados.*

### PREGUNTA 22

En investigación etiológica, la selección de un grupo de sujetos que padece una enfermedad y de un grupo de individuos que no la padece caracteriza a los estudios:

- A) De cohortes.
- B) Series de casos.
- C) Ensayos clínicos.
- D) De casos y controles.
- E) Transversales.

*Respuesta: D. Existen diferentes criterios para clasificar los diversos tipos de diseño de los estudios epidemiológicos analíticos. Uno de ellos es el criterio por el que se seleccionan los sujetos que conforman los grupos de estudio. Así, mientras que los estudios de cohortes se caracterizan porque los grupos se seleccionan según la presencia o la ausencia (o nivel de intensidad) de la exposición a un factor de riesgo, en los estudios de casos y controles se seleccionan según la presencia de una enfermedad (casos) o su ausencia (controles).*

### PREGUNTA 23

¿Cuál de los siguientes diseños es el más adecuado para estudiar la asociación entre una enfermedad y una exposición poco frecuente?

- A) Estudios de casos y controles.
- B) Estudios de cohortes.
- C) Estudios transversales.
- D) Estudios antes-después.
- E) Series de casos.

*Respuesta: A. Para estudiar la asociación entre una enfermedad y un hipotético factor de riesgo, debe utilizarse un diseño analítico. Los estudios que implican seguimiento de los sujetos no son buenos diseños para evaluar factores etiológicos poco frecuentes. Por el contrario, una de las principales ventajas de los estudios de casos y controles es que permiten evaluar exposiciones poco frecuentes en la población general, siempre y cuando estén fuertemente asociadas a la enfermedad en estudio.*

**PREGUNTA 24**

¿Cuál es la medida de asociación que se utiliza en los estudios de casos y controles?

- A) Riesgo relativo.
- B) Riesgo atribuible.
- C) Diferencia de incidencias.
- D) Cociente entre incidencias.
- E) Odds ratio.

*Respuesta: E. La medida epidemiológica de asociación más utilizada entre una exposición y una enfermedad es el riesgo relativo, que corresponde al cociente entre las incidencias en el grupo expuesto y el no expuesto. Sin embargo, en los estudios de casos y controles no puede determinarse la incidencia, puesto que no existe seguimiento de los sujetos, por lo que se utiliza una medida de asociación diferente, denominada odds ratio, que, en ciertas circunstancias, es una buena estimación del riesgo relativo.*

**PREGUNTA 25**

En un estudio se incluyen 100 casos, 60 de los cuales han estado expuestos al factor de riesgo, y 100 controles, de los cuales sólo 40 lo han estado. ¿Cuál es el valor de la odds ratio?

- A) 0,4.
- B) 0,6.
- C) 1.
- D) 2,25.
- E) 3.

*Respuesta: D. La odds ratio (OR) es la medida de asociación entre la enfermedad y la exposición en los estudios de casos y controles. Con los datos del enunciado de la pregunta anterior puede construirse la siguiente tabla 2 × 2:*

	Casos	Controles
Expuestos	60	40
No expuestos	40	60
Total	100	100

$$OR = 60 \times 60 / 40 \times 40 = 2,25$$

*Corresponde al cociente entre la odds de exposición entre los casos ( $60/40=1,5$ ) y la odds de exposición entre los controles ( $40/60=0,66$ ).*

**PREGUNTA 26**

¿Cómo clasificaría un estudio en el que los participantes reciben de forma aleatoria una determinada actividad preventiva o los cuidados habituales, con el objetivo de valorar la eficacia de esta actividad preventiva?

- A) Ensayo clínico.
- B) Estudio de cohortes.
- C) Estudio de casos y controles.
- D) Estudio de casos y controles dentro de una cohorte.
- E) Estudio de asociación cruzada.

*Respuesta: A. El ensayo clínico se define como aquel estudio en el que los investigadores controlan la intervención, y los participantes reciben de forma aleatoria la intervención de estudio o la de comparación.*

**PREGUNTA 27**

¿Cuál de las siguientes afirmaciones sobre los estudios de casos y controles es falsa?

- A) Son útiles para estudiar enfermedades poco frecuentes.

## Métodos de investigación clínica y epidemiológica

- B) Permiten estimar la incidencia de la enfermedad.
- C) Son muy vulnerables a la introducción de sesgos.
- D) Son útiles para estudiar enfermedades con largos períodos de latencia.
- E) Permiten evaluar más de un factor de riesgo simultáneamente.

*Respuesta: B. Los estudios de casos y controles son un diseño epidemiológico que se utiliza cada vez con mayor frecuencia, debido a que son muy eficientes al no requerir el seguimiento de los sujetos. Al mismo tiempo, presentan otras ventajas, como el hecho de que permiten estudiar enfermedades poco frecuentes y con largos períodos de latencia, puesto que se parte de los casos ya diagnosticados, y evaluar varios factores de riesgo de la enfermedad simultáneamente, analizando interacciones entre ellos. Sin embargo, presentan algunos inconvenientes derivados de su propio diseño, ya que son muy vulnerables a la introducción de sesgos y no permiten estimar directamente la incidencia de la enfermedad.*

### PREGUNTA 28

¿Qué tipo de estudios es el más adecuado para evaluar la eficacia de las intervenciones terapéuticas o preventivas?

- A) Estudios experimentales.
- B) Estudios observacionales.
- C) Estudios transversales.
- D) Estudios analíticos.
- E) Estudios prospectivos.

*Respuesta: A. La evaluación de la eficacia de las intervenciones terapéuticas y preventivas es fundamental para la medicina, especialmente con la aparición de gran cantidad de nuevas técnicas y procedimientos. Los estudios experimentales son los más adecuados para evaluar esta eficacia, ya que son los que proporcionan una mejor evidencia al permitir que los investigadores controlen («manipulen») la intervención. Aunque los estudios experimentales son prospectivos y analíticos, no todos los estudios prospectivos o analíticos son experimentales.*

### PREGUNTA 29

¿Cuál de las siguientes *no* es una restricción ética general de los estudios experimentales?

- A) Los sujetos deben ser informados de que participan en un experimento.
- B) Los sujetos no deben ser privados de las mejores medidas terapéuticas.
- C) Los sujetos deben conocer qué intervención están recibiendo.
- D) Las intervenciones en todos los grupos han de ser igualmente aceptables.
- E) Sólo pueden utilizarse con intervenciones potencialmente beneficiosas.

*Respuesta: C. Los estudios experimentales son los que proporcionan la mejor evidencia acerca de la relación entre un factor de estudio y un efecto. Sin embargo, no siempre es factible su realización, a menudo debido a las restricciones éticas de toda investigación en seres humanos. Básicamente, sólo deben ensayarse intervenciones potencialmente beneficiosas, informando a los sujetos de que participan en un estudio y de sus consecuencias, y no privándolos en ningún momento de las mejores medidas terapéuticas o preventivas que podrían recibir si no participaran en el estudio. Al mismo tiempo, sólo deben compararse intervenciones que sean igualmente aceptables para las características de los sujetos. Sin embargo, no es necesario informar de la intervención que recibe cada uno. De hecho, el enmascaramiento es una técnica utilizada habitualmente en este tipo de estudios.*

### PREGUNTA 30

Al evaluar la eficacia de una intervención, la diferencia principal entre un estudio experimental y uno observacional es que:

- A) El primero es prospectivo.
- B) El primero es retrospectivo.
- C) Los grupos de estudio son de igual tamaño.
- D) Los investigadores deciden quién recibe la intervención de estudio.
- E) El primero puede generalizar mejor los resultados.

*Respuesta: D. La característica principal que diferencia un estudio experimental de uno observacional es que, en el primero, existe manipulación del factor de estudio, es decir, que el equipo investigador decide quién recibe*

la intervención de estudio y quién no. Sin embargo, ciertos factores no pueden ser manipulados, ya sea por imposibilidad o por las restricciones éticas, y se recurre a diseños observacionales. Los estudios experimentales son prospectivos, pero también pueden serlo los observacionales.

### PREGUNTA 31

¿Cuál de las siguientes condiciones es imprescindible para poder clasificar un estudio como un ensayo clínico aleatorio?

- A) Que se estudie una muestra aleatoria de la población.
- B) Que se utilicen técnicas de enmascaramiento.
- C) Que la asignación de los sujetos a los grupos se realice al azar.
- D) Que pueda descartarse la existencia de un error aleatorio.
- E) Que el grupo control reciba placebo.

*Respuesta: C. El ensayo clínico aleatorio es el estudio experimental utilizado con mayor frecuencia para evaluar la eficacia de una intervención, comparándola con la de un grupo control, que a menudo recibe un placebo. Este tipo de diseño se caracteriza porque los investigadores manipulan el factor de estudio (como en cualquier estudio experimental) y porque la asignación de los sujetos a los grupos de estudio se realiza al azar (asignación aleatoria).*

### PREGUNTA 32

¿Cuál de las siguientes afirmaciones sobre la asignación aleatoria es cierta?

- A) Tiende a conseguir grupos comparables.
- B) Tiende a asegurar que la muestra es representativa.
- C) Evita el efecto placebo.
- D) Evita las pérdidas de seguimiento.
- E) Asegura que los grupos sean del mismo tamaño.

*Respuesta: A. La asignación aleatoria de los sujetos a los grupos de estudio es la característica definitoria de los ensayos clínicos aleatorios. Consiste en distribuir al azar a los sujetos entre los grupos que se van a comparar, con el fin de prevenir cualquier interferencia que el conocimiento de la intervención que se va a recibir pueda producir sobre el proceso de la asignación. De esta forma, se tiende a conseguir que los grupos sean comparables por todas las variables, tanto conocidas como desconocidas, en particular cuando el número de sujetos es elevado. No asegura necesariamente que el mismo número de sujetos sea asignado a cada grupo.*

### PREGUNTA 33

¿Cuál de las siguientes afirmaciones sobre la técnica de enmascaramiento denominada «doble ciego» es cierta?

- A) Evita el efecto placebo.
- B) Previene las pérdidas de seguimiento.
- C) Previene sesgos de información.
- D) Asegura que los grupos son comparables al inicio del estudio.
- E) Controla factores de confusión.

*Respuesta: C. En un ensayo clínico doble ciego, tanto el sujeto como el observador ignoran el tratamiento administrado. Las técnicas de enmascaramiento previenen sesgos de información pero no de selección, ni la presencia de factores de confusión.*

### PREGUNTA 34

¿Cuál de las siguientes afirmaciones sobre los ensayos cruzados es falsa?

- A) Tienen una duración menor que los estudios en paralelo.
- B) Cada participante actúa como su propio control.
- C) En el análisis pueden aplicarse pruebas estadísticas para datos apareados.
- D) Debe existir un período de blanqueo entre las dos fases del estudio.
- E) Requieren un menor número de sujetos que los estudios en paralelo.

*Respuesta: A. Los ensayos clínicos cruzados pueden tener un diseño muy eficiente en determinadas circunstancias. Los sujetos reciben dos tratamientos en dos períodos diferentes, de forma que debe existir un período*

## Métodos de investigación clínica y epidemiológica

de blanqueo entre ambas fases, que asegure que ha desaparecido el efecto del tratamiento administrado en primer lugar. Dado que cada sujeto actúa como su propio control, es necesario un número menor de participantes y pueden utilizarse pruebas estadísticas para datos apareados. Sin embargo, dado que cada sujeto debe recibir ambas intervenciones, la duración del estudio es mayor que en los diseños en paralelo.

### PREGUNTA 35

¿Cuál de las siguientes afirmaciones sobre los estudios «antes-después» es falsa?

- A) No permiten aislar el efecto del factor del estudio.
- B) No existe asignación aleatoria de los sujetos.
- C) Permiten comparar la eficacia de dos intervenciones.
- D) Carecen de grupo control.
- E) Cada sujeto actúa como su grupo control.

*Respuesta: C. Los estudios «antes-después» corresponden a un tipo de intervención en el que sólo existe un grupo de estudio. Se administra una intervención a un grupo de sujetos y se compara la situación al final del estudio con la que existía al inicio de éste, por lo que cada sujeto actúa como su propio control. Sin embargo, la ausencia de grupo control no permite aislar el efecto debido al factor de estudio del debido a otros factores, por lo que los resultados han de interpretarse con mucha precaución. Dado que sólo hay un grupo, no es posible comparar la eficacia de diferentes intervenciones.*

### PREGUNTA 36

¿Cuál de las siguientes afirmaciones sobre las revisiones sistemáticas de la evidencia científica es cierta?

- A) Corresponde a un estudio experimental.
- B) Siempre incluye una combinación estadística de los resultados de los estudios (metaanálisis).
- C) La obtención de un resultado global es especialmente útil cuando existe mucha heterogeneidad entre los estudios disponibles.
- D) Corresponde a un estudio observacional.
- E) Permite evitar el sesgo de publicación.

*Respuesta: D. Las revisiones sistemáticas de la evidencia científica pueden entenderse como estudios observacionales en los que la población de estudio está formada por las investigaciones realizadas sobre un tema concreto y accesibles a los revisores, por lo que no se evita el sesgo de publicación. Aunque es conveniente realizar una combinación estadística de los resultados de los estudios, ello no es siempre posible, ni es deseable cuando existe heterogeneidad entre los estudios.*

### PREGUNTA 37

¿Cuál de las siguientes afirmaciones sobre el análisis de heterogeneidad en una revisión sistemática es falsa?

- A) Si se detecta heterogeneidad, no es adecuado combinar los resultados de los estudios individuales.
- B) Si no se detecta heterogeneidad estadísticamente significativa, puede descartarse su existencia.
- C) El análisis de heterogeneidad puede utilizar métodos gráficos.
- D) Si se detecta heterogeneidad, deben investigarse sus posibles causas.
- E) El análisis de subgrupos puede ser ayuda en la valoración de las diferencias entre estudios.

*Respuesta: B. El análisis de heterogeneidad es clave en una revisión sistemática, y puede utilizar métodos gráficos y pruebas estadísticas. Sin embargo, las pruebas estadísticas pueden tener poca potencia, dado el habitualmente reducido número de estudios que se incluyen, por lo que no permiten descartar su existencia con seguridad. Si se detecta heterogeneidad, deben investigarse sus posibles causas y abstenerse de combinar los resultados de los estudios individuales. En la valoración de las diferencias entre estudios pueden ser de ayuda los análisis de subgrupos.*

### PREGUNTA 38

¿Cuál de las siguientes afirmaciones sobre el análisis y la presentación de resultados en una revisión sistemática es falsa?

- A) El cálculo estadístico de una medida global a partir de los resultados de los estudios individuales se denomina metaanálisis.

- B) La combinación estadística de los resultados de los estudios individuales sólo está justificada si se ha descartado la existencia de heterogeneidad.
- C) El análisis de sensibilidad permite determinar la consistencia del resultado independientemente de los criterios utilizados en la selección de los estudios.
- D) Una ventaja de la revisión sistemática es que permite realizar análisis de subgrupos aunque no se hayan planeado a priori.
- E) Los resultados de los estudios individuales se expresan de forma estandarizada para permitir su comparación.

*Respuesta: D. Una de las ventajas de las revisiones sistemáticas es que permiten realizar análisis de subgrupos con mayor potencia estadística que los estudios individuales y, en ocasiones, sirven para contrastar hipótesis que no podrían abordarse en un único estudio. Sin embargo, están sujetas a los mismos criterios de validez que cualquier análisis de subgrupos, incluyendo que deben haber sido planeadas a priori.*

### PREGUNTA 39

¿Cuál de los siguientes no es un criterio de validez de un análisis de subgrupos?

- A) Respuesta a una hipótesis planteada a priori.
- B) Plausibilidad biológica.
- C) Escasa magnitud del efecto.
- D) Nivel de significación estadística más estricto.
- E) Comparaciones realizadas dentro de un mismo estudio.

*Respuesta: C. Los análisis de subgrupos se utilizan con mucha frecuencia para evaluar el efecto de una intervención en subgrupos de sujetos o estratos definidos por otras características. Para que sean válidos, deben responder a hipótesis planteadas a priori, plausibles biológicamente y realizadas en una estrategia de comparaciones múltiples. Se considera que tienen mayor validez si encuentran resultados de gran magnitud, estadísticamente muy significativos (nivel de significación más estricto) y si responden a comparaciones realizadas dentro de un mismo estudio, y no a comparaciones entre subgrupos de estudios diferentes.*

### PREGUNTA 40

En la evaluación de la validez de un estudio, un criterio fundamental es la comparabilidad inicial de los grupos. Si existe alguna diferencia en el pronóstico inicial de los grupos, ¿qué afirmación es cierta?

- A) El estudio carece de validez interna.
- B) El análisis estadístico debe realizarse ajustando por los potenciales factores de confusión.
- C) No debe tenerse en cuenta si el tamaño de la muestra es elevado.
- D) Sólo invalida el estudio si corresponde a un diseño de cohortes.
- E) Invalida el estudio si se encuentra una diferencia de relevancia clínica.

*Respuesta: B. La comparabilidad inicial por los factores pronósticos es un importante criterio de validez de los estudios. Sin embargo, si se detecta alguna diferencia que se considera relevante y que puede influir sobre los resultados, debe tenerse en cuenta en el análisis, ajustando la estimación del efecto o la asociación por los potenciales factores de confusión, ya sea mediante un análisis estratificado o una técnica multivariante.*

### PREGUNTA 41

En la lectura crítica de un ensayo clínico, ¿cuál de los siguientes aspectos no se refiere a la evaluación de su validez interna?

- A) La representatividad de la muestra de sujetos estudiados respecto a la población que presenta la enfermedad.
- B) La estrategia de análisis por intención de tratar.
- C) La asignación aleatoria de los sujetos a los grupos de estudio.
- D) El seguimiento completo de todos los sujetos.
- E) La evaluación de la comparabilidad al inicio del estudio.

*Respuesta: A. En la evaluación de la validez de un ensayo clínico, deben tenerse en cuenta los aspectos que se relacionan con la comparabilidad de los grupos, tanto al inicio del estudio (asignación aleatoria y oculta a los investigadores, estrategia de análisis por intención de tratar, evaluación de la comparabilidad inicial por factores pronósticos) como a lo largo del mismo (utilización de técnicas de enmascaramiento, evaluación a ciegas de la*

## Métodos de investigación clínica y epidemiológica

respuesta, seguimiento completo de los sujetos). Los aspectos relacionados con las características de los pacientes tienen más que ver con la aplicabilidad práctica de los resultados que con la validez interna del estudio.

### PREGUNTA 42

Buscando evidencias sobre la eficacia de la vacunación antineumocócica, localiza un ensayo clínico aleatorio que la compara con placebo en sujetos sanos. Se observa una incidencia de neumonía neumocócica en el 3% de los sujetos que han recibido placebo y en el 1% de los que han sido vacunados. ¿Cuál es la reducción absoluta del riesgo de neumonía neumocócica observada en este estudio?

- A) 1%.
- B) 2%.
- C) 3%.
- D) 3.
- E) 1/3 (33%).

**Respuesta:** B. La medida absoluta de la reducción de riesgo conseguida con la vacuna en relación con el placebo corresponde a la diferencia entre las incidencias de neumonía neumocócica observadas en cada grupo:  $3\%-1\% = 2\%$ . Indica que la vacuna antineumocócica ha producido una reducción del 2% en la incidencia de neumonía neumocócica en los sujetos estudiados en relación con el placebo.

### PREGUNTA 43

En el estudio de la pregunta anterior, ¿cuál es la reducción relativa del riesgo de neumonía neumocócica conseguido con la vacuna?

- A) 1/3 (33%).
- B) 3.
- C) 2/3 (67%).
- D) 2%.
- E) 2.

**Respuesta:** C. La medida relativa de la reducción del riesgo de neumonía neumocócica conseguida con la vacuna en relación con el placebo corresponde a la proporción que la diferencia entre las incidencias de neumonía observadas en cada grupo supone en relación con la incidencia del grupo de referencia (placebo):  $(3\%-1\%)/3\% = 2/3$  (67%). Indica que la vacuna ha reducido en dos terceras partes la incidencia de la neumonía neumocócica en relación con el placebo. Expresado en términos de riesgo relativo (RR), se obtiene  $RR = 1/3$ , es decir, la indicencia de neumonía en el grupo vacunado es la tercera parte que en el grupo placebo (es decir, se ha reducido la incidencia en dos terceras partes).

### PREGUNTA 44

En el estudio de la pregunta anterior, ¿cuál es el número de personas que es necesario vacunar para evitar un caso adicional de neumonía neumocócica en los sujetos estudiados?

- A) 100.
- B) 50.
- C) 33.
- D) 3.
- E) 10.

**Respuesta:** B. La reducción absoluta de riesgo de neumonía de la vacuna en relación con placebo observada en los sujetos del estudio es  $3\%-1\% = 2\%$ . Es decir, por cada 100 pacientes vacunados se evitan dos casos de neumonía neumocócica más que con placebo, por lo que el valor del NNT es  $100/2 = 50$ . Es decir, es necesario vacunar a 50 sujetos similares a los del estudio para evitar un caso de neumonía neumocócica más que si se les administrara placebo. El NNT corresponde al inverso de la reducción absoluta de riesgo:  $1/0,02 = 50$ .

### PREGUNTA 45

Suponga que desea aplicar los resultados del estudio de la pregunta anterior a un paciente de su consulta que usted considera que tiene un mayor riesgo de contraer una neumonía

neumocócica que el promedio de los sujetos del ensayo, estimando que su probabilidad de hacerlo es aproximadamente del doble de la observada en el estudio. ¿Cuántos pacientes como éste sería necesario vacunar para evitar un caso de neumonía más que con placebo?

- A) 100.
- B) 200.
- C) 50.
- D) 25.
- E) 133.

*Respuesta: D. Si el sujeto tiene el doble de riesgo que los del estudio, el NNT será aproximadamente la mitad del observado en el mismo, es decir,  $50/2=25$ . El riesgo basal del sujeto de interés es el doble del observado en el grupo control en el estudio, es decir, del 6%. Asumiendo que la eficacia de la vacuna no varía, es decir, que reduce el riesgo a una tercera parte, puede decirse que se espera que reduzca la incidencia a un 2%. Por tanto, la reducción absoluta de riesgo que se produciría es del  $6\%-2\% = 4\%$ , y el NNT en este caso sería  $100/4=25$ . Es decir, sería necesario vacunar a 25 sujetos similares al de interés para evitar un caso de neumonía neumocócica más que si les administrara placebo.*

## PREGUNTAS CON RESPUESTA VERDADERO/FALSO

### PREGUNTA 1

En un estudio transversal, los datos de cada sujeto representan un momento del tiempo.

**Respuesta:** Verdadero. Se consideran transversales los estudios en los que los datos de cada sujeto representan esencialmente un momento del tiempo. Estos datos pueden corresponder a la presencia, a la ausencia o a diferentes grados de una característica o enfermedad (como ocurre, por ejemplo, en los estudios de prevalencia de un problema de salud en una comunidad determinada), o bien examinar la relación.

### PREGUNTA 2

En un estudio longitudinal se puede establecer una secuencia temporal entre las variables del estudio.

**Respuesta:** Verdadero. Se consideran longitudinales los estudios en los que existe un lapso de tiempo entre las distintas variables que se evalúan, de forma que puede establecerse una secuencia temporal entre ellas. Pueden ser tanto descriptivos como analíticos.

### PREGUNTA 3

En un estudio de casos y controles, la dirección temporal va de la causa al desenlace.

**Respuesta:** Falso. Se consideran longitudinales los estudios en los que existe un lapso de tiempo entre las distintas variables que se evalúan, de forma que puede establecerse una secuencia temporal entre ellas. Pueden ser tanto descriptivos como analíticos. En estos últimos debe tenerse en cuenta, además, la dirección temporal, que puede ir de la causa hacia el desenlace (estudios experimentales y estudios de cohortes) o bien desde el desenlace hacia la causa (estudios de casos y controles).

### PREGUNTA 4

En un estudio observacional, los investigadores controlan el factor de estudio pero no hay asignación aleatoria.

**Respuesta:** Falso. Se definen como observacionales los estudios en los que el factor de estudio no es controlado por los investigadores, sino que éstos se limitan a observar, medir y analizar determinadas variables en los sujetos. La exposición puede venir impuesta (p. ej., el sexo o la raza), haber sido «escogida» por los propios sujetos (p. ej., el consumo de tabaco), o decidida por el profesional sanitario dentro del proceso habitual de atención sanitaria (p. ej., los actos terapéuticos ordinarios), pero no de forma deliberada en el marco de una investigación. Se consideran experimentales los estudios en los que el equipo investigador asigna el factor de estudio y lo controla de forma deliberada para la realización de la investigación, según un plan preestablecido. Estos estudios se centran en una relación causa-efecto (analíticos), y en general evalúan el efecto de una o más intervenciones preventivas o terapéuticas.

### PREGUNTA 5

Un ensayo clínico es un estudio prospectivo.

**Respuesta:** Verdadero. El ensayo clínico aleatorio (ECA) en su forma más sencilla es un estudio de diseño paralelo con dos grupos, que consiste en la selección de una muestra de pacientes y su asignación de forma aleatoria a uno de los dos grupos. Uno de ellos recibe la intervención de estudio y el otro la de control, que se utiliza como referencia o comparación. Ambos grupos se siguen de forma concurrente durante un período determinado, cuantificando y comparando las respuestas observadas en ambos.

### PREGUNTA 6

Un ensayo clínico es siempre un estudio analítico.

**Respuesta:** Verdadero. El objetivo básico de los estudios experimentales es evaluar los efectos de una intervención, intentando establecer una relación causa-efecto con los resultados observados (son, pues, estudios analíticos).

### PREGUNTA 7

Un estudio analítico es siempre el mejor estudio para demostrar la eficacia de una intervención.

**Respuesta:** Falso. El mejor tipo de estudio es el ensayo clínico, que es un estudio analítico prospectivo. Sin embargo, los estudios de cohortes también son analíticos y pueden ser prospectivos, pero no son la mejor elección para estimar la eficacia de un tratamiento.

**PREGUNTA 8**

La principal diferencia entre un ensayo clínico y un estudio de cohortes es la asignación aleatoria de la intervención.

**Respuesta:** Verdadero. *En los ensayos clínicos existe asignación aleatoria, pero no en los estudios de cohortes.*

**PREGUNTA 9**

La principal diferencia entre un estudio experimental y un estudio de cohortes es la asignación aleatoria.

**Respuesta:** Falso. *La principal diferencia es que en los estudios experimentales se ha contado el factor de estudio, mientras que los de cohortes son estudios observacionales sin control del factor de estudio. En ocasiones, en los estudios experimentales los sujetos son su propio control y no existe asignación aleatoria. En los estudios de cohortes nunca existe asignación aleatoria.*

**PREGUNTA 10**

La finalidad del placebo es controlar el efecto Hawthorne.

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

**PREGUNTA 11**

El uso del placebo permite enmascarar las intervenciones.

**Respuesta:** Verdadero. *El uso de un placebo permite enmascarar las intervenciones, de forma que los participantes (e incluso los investigadores) desconozcan si reciben tratamiento activo o no.*

**PREGUNTA 12**

En un ensayo clínico, la exclusión de los sujetos incumplidores antes de la asignación aleatoria aumenta la potencia del estudio.

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

**PREGUNTA 13**

En un ensayo clínico, la exclusión de los sujetos incumplidores antes de la asignación aleatoria aumenta la potencia del estudio y su validez externa.

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

**PREGUNTA 14**

La asignación aleatoria asegura siempre la comparabilidad de los grupos en un ensayo clínico.

**Respuesta:** Falso. *La asignación aleatoria tiende a asegurar la comparabilidad de los grupos, pero no siempre se consigue. Esta tendencia es tanto mayor cuanto más elevado es el tamaño de la muestra. Cuando se estudian pocos pacientes pueden producirse diferencias entre los grupos, a pesar de que los sujetos se hayan asignado de forma aleatoria.*

## Métodos de investigación clínica y epidemiológica

### PREGUNTA 15

Cuanto mayor es el tamaño de la muestra en un ensayo clínico, mayor es la probabilidad de que los grupos sean comparables.

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

### PREGUNTA 16

El uso de técnicas de enmascaramiento es muy útil cuando la variable de respuesta es subjetiva.

**Respuesta:** Verdadero. Estas técnicas son muy importantes cuando la variable de respuesta es blanda e incluye algún elemento de subjetividad (p. ej., una medida de la intensidad del dolor, un cuestionario sobre síntomas, la percepción de mejoría en relación con una situación previa o la interpretación de una radiografía). En cambio, cuando la variable de respuesta es objetiva (dura), como una medida de laboratorio, esta técnica no es imprescindible.

### PREGUNTA 17

Si se quiere mantener intacta la comparabilidad conseguida con la asignación aleatoria, la mejor alternativa es evaluar a todos los participantes según el principio conocido como análisis de casos válidos o por protocolo.

**Respuesta:** Falso. Si se quiere mantener intacta la comparabilidad conseguida con la asignación aleatoria, la mejor alternativa es evaluar a todos los pacientes según el principio conocido como análisis por intención de tratar o según la asignación aleatoria, es decir, manteniendo a cada sujeto en el grupo al que ha sido asignado, independientemente de si ha cumplido o no con la intervención asignada o de si ha recibido otro tratamiento diferente del asignado. Esta estrategia de análisis se considera la única válida en un ensayo clínico, ya que, además de mantener la comparabilidad de los grupos, es la más conservadora, en el sentido de que es la opción con la que resulta más difícil obtener resultados positivos favorables a la nueva intervención. Otra opción de análisis es considerar exclusivamente los sujetos de cada uno de los grupos que han cumplido el protocolo del ensayo, han recibido el tratamiento asignado y han finalizado el seguimiento del estudio. Esta estrategia se denomina análisis de casos válidos o por protocolo.

### PREGUNTA 18

Una actitud pragmática en un ensayo clínico consiste en establecer criterios de selección muy estrictos que definen una población muy homogénea.

**Respuesta:** Falso. La actitud explicativa consiste en establecer criterios de selección muy estrictos, que definen una población muy homogénea, con escasa variabilidad, buena cumplidora, en la que sea más fácil obtener datos de calidad y en la que exista una mayor probabilidad de encontrar un efecto o asociación si existe. La actitud pragmática consiste en establecer unos criterios de selección amplios, que definen una población heterogénea, más representativa de la población general. Su inconveniente es que se pierde cierto grado de control sobre la situación y puede diluir o enmascarar una asociación o un efecto existentes, por lo que se complica la interpretación de los resultados.

### PREGUNTA 19

Los ensayos clínicos con asignación de grupos son útiles para evitar la contaminación que puede producirse entre las intervenciones.

**Respuesta:** Verdadero. Habitualmente, en un ECA cada sujeto se asigna de forma individual a uno de los grupos que se comparan, recibe directamente la intervención asignada y sobre él se mide la respuesta observada. Sin embargo, en ocasiones, esta asignación individual no es posible o no es adecuada, por lo que se

recurre a una asignación por grupos (clusters). Este diseño es útil cuando se pretende evitar la posible contaminación que puede producirse entre las intervenciones al realizarse en un mismo emplazamiento.

## PREGUNTA 20

El diseño factorial de un ensayo clínico permite evaluar dos o más intervenciones en un único estudio.

**Respuesta:** Verdadero. El diseño factorial permite al investigador evaluar dos intervenciones o más en un único estudio. En su forma más sencilla para evaluar dos tratamientos, cada sujeto se asigna aleatoriamente a uno de cuatro grupos: un grupo recibe ambos tratamientos, dos grupos reciben solamente uno de los tratamientos cada uno, y el cuarto grupo no recibe ningún tratamiento. Si existe enmascaramiento, cada uno de los tres últimos grupos debe recibir un placebo del tratamiento o de los tratamientos que no se administran.

## PREGUNTA 21

En un ensayo clínico cruzado, cuando la respuesta en el segundo período está afectada por el tratamiento recibido durante el primero, se produce un efecto período.

**Respuesta:** Falso. Dado que cada paciente se observa en dos períodos distintos, es importante determinar si ha existido algún cambio entre el primero y el segundo. La enfermedad o el síntoma puede progresar, regresar o fluctuar en su gravedad, por lo que es posible encontrar cambios entre los diferentes períodos, con independencia del tratamiento administrado. Una manera de minimizar este efecto es que cada paciente cambie de tratamiento varias veces, recibiéndolos durante varios intervalos de tiempo, de forma que si existe un efecto período, éste quede contrarrestado. Sin embargo, ello dificulta mucho la ejecución del estudio y también el que los pacientes acepten participar. Un efecto secuencia puede darse siempre que un individuo se observa más de una vez en períodos distintos.

## PREGUNTA 22

En los ensayos clínicos cruzados, al evaluar los dos tratamientos en un mismo paciente, la variabilidad intraindividual es menor.

**Respuesta:** Verdadero. Al evaluar los dos tratamientos en un mismo paciente, la variabilidad intraindividual es menor, lo que permite utilizar pruebas estadísticas para datos apareados que tienen una mayor potencia y, por tanto, se precisa un número todavía menor de sujetos.

## PREGUNTA 23

Los ensayos clínicos cruzados suelen tener más potencia estadística que los ensayos clínicos en paralelo.

**Respuesta:** Verdadero. Al evaluar los dos tratamientos en un mismo paciente, la variabilidad intraindividual es menor, lo que permite utilizar pruebas estadísticas para datos apareados que tienen una mayor potencia y, por tanto, se precisa un número todavía menor de sujetos.

## PREGUNTA 24

Los ensayos clínicos cruzados son muy útiles para evaluar la eficacia de un tratamiento curativo para una enfermedad aguda.

**Respuesta:** Falso. El diseño cruzado no deberá utilizarse en enfermedades agudas o que cursen a brotes, ni cuando el orden en que se administren las intervenciones pueda alterar el resultado, o cuando no sea posible realizar un período de blanqueo que asegure la total desaparición del efecto de la primera intervención.

## PREGUNTA 25

Un estudio de cohortes es un diseño transversal observacional en el que se comparan dos o más grupos.

**Respuesta:** Falso. Un estudio de cohortes es un diseño observacional analítico longitudinal en el que se comparan dos cohortes o dos grupos dentro de una misma cohorte, que difieren por su exposición al factor de estudio, con el objetivo de evaluar una posible relación causa-efecto. Cuando solamente existe una cohorte que

## Métodos de investigación clínica y epidemiológica

es seguida en el tiempo con la única finalidad de estimar la incidencia con que aparece un determinado problema de salud (desenlace o efecto) o describir su evolución, se trata de un diseño descriptivo longitudinal.

### PREGUNTA 26

El carácter prospectivo de un estudio de cohortes indica que el investigador recoge la información sobre la exposición cuando se inicia el estudio.

**Respuesta:** Verdadero. El carácter prospectivo significa que el investigador recoge la información sobre la exposición cuando se inicia el estudio, e identifica los nuevos casos de la enfermedad o las defunciones que se producen a partir de ese momento, es decir, el inicio del estudio es anterior al desarrollo de los hechos y desenlaces.

### PREGUNTA 27

En los estudios de cohortes, a diferencia de lo que ocurre en los ensayos clínicos, no se pueden calcular las tasas de incidencia y el riesgo relativo.

**Respuesta:** Falso. En los estudios de cohortes se pueden calcular las tasas de incidencia y el riesgo relativo, lo que supone una ventaja en relación con los estudios de casos y controles, que tan sólo permiten estimar el riesgo relativo de forma aproximada pero no calcular las tasas de incidencia.

### PREGUNTA 28

Los estudios de cohortes permiten evaluar los efectos del factor de riesgo sobre varias enfermedades.

**Respuesta:** Verdadero. Los estudios de cohortes permiten evaluar los efectos del factor de riesgo sobre varias enfermedades, por ejemplo de la hipertensión arterial en relación con diferentes desenlaces cardiovasculares.

### PREGUNTA 29

Los estudios de cohortes son eficientes para el estudio de enfermedades poco frecuentes.

**Respuesta:** Falso. Los estudios de cohortes no son eficientes para el estudio de enfermedades poco frecuentes, ya que obligan a seguir a un número elevado de sujetos durante un largo período de tiempo para observar un número suficiente de casos de la enfermedad. De todos modos, aunque la frecuencia de una enfermedad sea muy baja en la población general, puede ser lo suficientemente elevada en grupos especiales que hayan estado expuestos de forma intensa a un factor de riesgo y, en consecuencia, puede llevarse a cabo un estudio de cohortes en este grupo de población expuesto.

### PREGUNTA 30

Los estudios de cohortes prospectivos no son eficientes para el estudio de enfermedades con un largo período de latencia.

**Respuesta:** Verdadero. Los estudios de cohortes prospectivos tampoco son eficientes para el estudio de enfermedades con un largo período de latencia, ya que obligan a un seguimiento muy largo, lo cual, además del coste y esfuerzo que supone, puede aumentar las pérdidas.

### PREGUNTA 31

Los estudios de cohortes retrospectivos permiten asegurar una adecuada secuencia temporal.

**Respuesta:** Verdadero. En los estudios retrospectivos de cohortes, también llamados estudios de cohortes históricas, el investigador identifica una cohorte de sujetos según sus características en el pasado, y reconstruye su experiencia de enfermedad hasta un momento definido de tiempo en el pasado más reciente o en el momento actual. La estructura de un estudio retrospectivo de cohortes es similar a la de un estudio prospectivo: se efectúa un seguimiento de un grupo de individuos a lo largo del tiempo, en los que se miden las posibles variables predictoras al inicio y después se determinan los desenlaces que se producen. Por tanto, ambos diseños permiten asegurar que la exposición (factor de riesgo) precede a la enfermedad o el efecto. La diferencia es que la identificación de la cohorte, las mediciones basales, el seguimiento y los desenlaces ya se han producido cuando se inicia el estudio.

**PREGUNTA 32**

En los estudios de casos y controles se elige un grupo de individuos que tienen una exposición determinada (casos), y otro en la que está ausente (controles).

**Respuesta:** Falso. En los estudios de casos y controles se elige un grupo de individuos que tienen un efecto o una enfermedad determinada (casos), y otro en el que está ausente (controles). Ambos grupos se comparan respecto a la frecuencia de exposición previa a un factor de riesgo (factor de estudio) que se sospecha que está relacionado con dicho efecto o enfermedad.

**PREGUNTA 33**

En los estudios de casos y controles, siempre que sea posible, es preferible incluir sólo casos prevalentes, que lleven un tiempo diagnosticados.

**Respuesta:** Falso. Un aspecto importante es el momento en la historia de la enfermedad en que se diagnostican los casos. Siempre que sea posible, es preferible incluir sólo casos recién diagnosticados (casos incidentes). Los casos prevalentes pueden ser los supervivientes, y la supervivencia puede estar relacionada con la exposición. En este caso, se hallaría un riesgo menor que al analizar casos incidentes. En el otro extremo, si la variable en estudio es un factor de buen pronóstico, la inclusión de casos prevalentes llevaría a la sobreestimación de la asociación.

**PREGUNTA 34**

En los estudios de casos y controles, los controles deben ser comparables a los casos en el sentido de haber tenido la misma probabilidad de exposición.

**Respuesta:** Verdadero. Los controles deben ser comparables a los casos en el sentido de haber tenido la misma probabilidad de exposición. En la práctica, hay que excluir del grupo control a los pacientes que tengan enfermedades relacionadas de forma positiva o negativa con la exposición en estudio. Si se quisiera estudiar el posible papel protector de la aspirina en la enfermedad coronaria, se debería excluir como controles, por ejemplo, a los pacientes con enfermedades reumáticas (pueden tomar aspirina como parte de su tratamiento) o aquellos con enfermedades gástricas (que probablemente evitarán su uso).

**PREGUNTA 35**

El sesgo de detección de los estudios de casos y controles es un sesgo de selección.

**Respuesta:** Verdadero. El llamado sesgo de detección aparece como consecuencia de la existencia de una diferente probabilidad de diagnosticar a los casos y a los controles. Este sesgo mereció la atención de muchos investigadores a raíz de la controversia surgida con el uso de estrógenos y el riesgo de padecer cáncer de endometrio. Se argumentó que la exposición (uso de estrógenos) estaba asociada a una mayor probabilidad de que los médicos detectaran la enfermedad, ya que las mujeres que utilizaban estrógenos eran visitadas con mayor frecuencia y de forma más cuidadosa que las que no los consumían. Sin embargo, la asociación entre el uso de estrógenos y el cáncer de endometrio permaneció después de considerar el posible efecto de un sesgo de detección.

**PREGUNTA 36**

El sesgo Berkson de los estudios de casos y controles es un sesgo de selección.

**Respuesta:** Verdadero. El sesgo de Berkson suele ocurrir cuando la combinación de la exposición y la enfermedad en estudio aumenta la probabilidad de ingreso en un hospital, lo cual conduce a una frecuencia de exposición sistemáticamente más elevada en los casos que en los controles hospitalarios. El resultado es una estimación sesgada de la odds ratio. Se trata de un sesgo de selección.

**PREGUNTA 37**

En los estudios de casos y controles, el uso de casos prevalentes en lugar de incidentes puede dar lugar a un error sistemático que se conoce con el nombre de sesgo de Neyman o de supervivencia selectiva.

**Respuesta:** Verdadero. El uso de casos prevalentes en lugar de incidentes puede dar lugar a un error sistemático que se conoce con el nombre de sesgo de Neyman o de supervivencia selectiva. La prevalencia depende tanto de la duración de la enfermedad, que se verá afectada por el tratamiento y la atención sanitaria recibida, como de la letalidad de la enfermedad. Por estas razones, los casos prevalentes pueden no ser representativos de todos los casos.

### PREGUNTA 38

Cuando se emparejan casos y controles por variables que no son factores de confusión se produce el fenómeno denominado sobreemparejamiento.

**Respuesta:** Verdadero. Para que el emparejamiento sea útil, debe realizarse por una variable que sea un potencial factor de confusión, es decir, relacionada tanto con el factor de estudio (exposición) como con la variable de respuesta (enfermedad). En caso de que no sea así, se pierde eficiencia y se disminuye la validez de la comparación entre los casos y los controles, fenómeno que se conoce como sobreemparejamiento (over-matching). Este fenómeno también se produce cuando el emparejamiento no se acompaña de una técnica de análisis para datos apareados.

### PREGUNTA 39

Los estudios de cohortes son más eficientes que los de casos y controles.

**Respuesta:** Falso. Los estudios de casos y controles son más eficientes, ya que se realizan en mucho menos tiempo y con un coste menor que los estudios de cohortes, al no tener que seguir a los pacientes durante el período de latencia de la enfermedad.

### PREGUNTA 40

Los estudios de cohortes, al elegir la población a partir de la exposición, están menos sujetos a la introducción de errores sistemáticos que los estudios de controles.

**Respuesta:** Verdadero. La limitación más importante es que son muy vulnerables a la introducción de sesgos, tanto en la selección de los grupos como en la información que se obtiene sobre los factores de riesgo. Los estudios de cohortes, al elegir la población a partir de la exposición, están menos sujetos a la introducción de errores sistemáticos.

### PREGUNTA 41

Los estudios de casos y controles no proporcionan una estimación directa de la incidencia ni de la prevalencia de una enfermedad.

**Respuesta:** Verdadero. Los estudios de casos y controles no proporcionan una estimación directa de la incidencia ni de la prevalencia de una enfermedad, ya que la proporción de los participantes en el estudio que tienen la exposición está determinada por el investigador, no por la proporción que existe en la comunidad.

### PREGUNTA 42

Los estudios de prevalencia son útiles para analizar la frecuencia de enfermedades raras.

**Respuesta:** Falso. La enfermedad estudiada debe ser de inicio lento y de larga duración, ya que, en caso contrario, existirán pocos individuos que la padecan en un momento determinado. Estos estudios no son útiles en el caso de enfermedades poco frecuentes, ya que para encontrar un número suficiente de casos debería recogerse información de una gran cantidad de personas.

### PREGUNTA 43

En los estudios de prevalencia, siempre que hay personas que no responden a los cuestionarios o no participan en el estudio se produce un sesgo.

**Respuesta:** Falso. Un sesgo se produce cuando la prevalencia entre los que responden o participan en el estudio es distinta de la que se observa entre los que no participan (no respuestas).

### PREGUNTA 44

En un estudio que evalúa una nueva prueba diagnóstica, si se estudiaran sólo aquellos pacientes que presentasen un proceso más grave, o más extenso, la nueva prueba daría más resultados positivos que si el grupo lo compusieran pacientes menos graves.

**Respuesta:** Verdadero. Los pacientes en los que se evalúa una nueva prueba no han de diferir sustancialmente de la población a la que se aplicará en la práctica clínica. Si se estudiaran sólo aquellos que presentasen un proceso más grave, o más extenso, la nueva prueba daría más resultados positivos que si el grupo lo compusieran

pacientes menos graves. Por esta razón, en el grupo de estudio se han de incluir pacientes con diferentes estados clínicos y patológicos de la enfermedad. Así mismo, es importante que no se excluyan aquellos sujetos que se encuentren en el límite diagnóstico.

### **PREGUNTA 45**

En un estudio que evalúa una nueva prueba diagnóstica, el grupo de no enfermos permite determinar los falsos positivos.

**Respuesta: Verdadero.** El grupo de no enfermos permite determinar los falsos positivos. En este grupo se deben incluir individuos que tengan enfermedades que planteen problemas de diagnóstico diferencial con los pacientes del otro grupo. Si se incluyen sólo individuos sanos, existe el peligro de encontrar un número menor de falsos positivos, lo que se traduce en una especificidad de la prueba más elevada.

### **PREGUNTA 46**

Si en un estudio que evalúa una prueba diagnóstica se parte de dos grupos, uno formado por pacientes con la enfermedad diagnosticada a partir del criterio de referencia y otro formado por individuos sin ella, no se puede evaluar el valor predictivo positivo.

**Respuesta: Verdadero.** Con esta estrategia, la prevalencia de la enfermedad en la población de estudio (la relación entre el grupo de enfermos y el de no enfermos) puede ser muy diferente de la que existe en la práctica. Así, por ejemplo, si se incluye el mismo número de enfermos que de no enfermos, la prevalencia en los sujetos estudiados es del 50%. Si bien permite la determinación de la sensibilidad, de la especificidad de forma eficiente, de los cocientes de probabilidad y de las curvas ROC, dificulta la estimación de los valores predictivos (anexo 3), ya que éstos dependen de la prevalencia de la enfermedad, y ésta depende a su vez del número de sujetos incluidos en cada grupo.

### **PREGUNTA 47**

La evaluación de la repetibilidad de una prueba diagnóstica requiere que dos observadores distintos analicen los resultados de la prueba.

**Respuesta: Falso.** La evaluación de la repetibilidad tiene por objetivo determinar si una prueba da los mismos resultados, o similares, cuando se aplica a una misma persona en más de una ocasión. Requiere que las condiciones de aplicación sean iguales en todas las ocasiones.

### **PREGUNTA 48**

Una revisión sistemática es un estudio de investigación que se realiza sobre datos primarios, es decir, utiliza los datos recogidos previamente en otros estudios.

**Respuesta: Falso.** Los datos que se utilizan en las revisiones sistemáticas son secundarios, es decir, se han recogido previamente en otros estudios.

### **PREGUNTA 49**

Los conceptos de revisión sistemática y metaanálisis son sinónimos.

**Respuesta: Falso.** Aunque con frecuencia los términos revisión sistemática y metaanálisis se emplean como sinónimos, hoy en día se reserva el concepto de metaanálisis para el procedimiento estadístico que se utiliza para integrar los resultados de diferentes estudios y obtener una estimación conjunta, mientras que el de revisión sistemática se aplica a la síntesis de la mejor evidencia disponible para responder a una pregunta concreta mediante la aplicación de una metodología explícita y rigurosa. Para ello, se utiliza la misma metodología que cualquier estudio de investigación: se formula un objetivo, se busca, se identifican y se valoran las pruebas disponibles siguiendo un protocolo con criterios explícitos, y se obtienen conclusiones a partir del análisis de los datos recogidos y la interpretación de los resultados. Cuando es posible se efectúa una síntesis estadística de los datos y entonces hablamos de metaanálisis.

### **PREGUNTA 50**

El sesgo de publicación en una revisión sistemática se refiere al hecho de que no todos los estudios tienen la misma probabilidad de ser publicados.

**Respuesta: Verdadero.** El sesgo de publicación se refiere al hecho de que no todos los estudios tienen la misma probabilidad de ser publicados. Con frecuencia, los autores deciden no enviar sus manuscritos a las

## Métodos de investigación clínica y epidemiológica

revistas, o bien los editores y los revisores de una revista deciden no aceptar algunos estudios en función de determinadas características, relacionadas más con los resultados encontrados que con aspectos de calidad.

### PREGUNTA 51

Un tesauro es un índice que lleva asociado un diccionario de sinónimos.

**Respuesta:** Falso. Un tesauro no es un índice ni un diccionario de sinónimos, sino un conjunto de términos que permiten expresar con la mayor exactitud posible una determinada noción o concepto. Por lo tanto, debe ser conciso, unívoco, bien estructurado, inteligible y capaz de evolucionar. Un tesauro es una forma de clasificación y de indización de una base de datos. Se basa en el uso de palabras clave que forman un vocabulario especializado cuyos términos se relacionan entre sí de forma jerárquica. Su finalidad es «traducir» el lenguaje natural empleado en los documentos a un vocabulario controlado de términos que permite representar el contenido de los documentos y facilita la indización y la consulta de la base de datos. Sólo los términos que figuran en una lista predefinida pueden ser utilizados. El hecho de que un concepto esté representado por un único término descriptor resuelve los problemas de sinonimias y polisemias.

### PREGUNTA 52

El MeSH (*Medical Subject Headings*) es el tesauro de la base de datos EMBASE.

**Respuesta:** Falso. El tesauro más utilizado en el ámbito de la medicina es el MeSH (*Medical Subject Headings*), que corresponde al vocabulario o tesauro preestablecido de descriptores, utilizado para indizar diferentes bases de datos de la United States National Library of Medicine (NLM). Utilizado correctamente, este vocabulario constituye una herramienta poderosa que mejora el acceso a la literatura biomédica. Refleja todo el ámbito de la literatura biomédica. La NLM lo mantiene, añadiéndole nuevos términos, a medida que aparecen nuevos conceptos, o bien modificando o eliminando otros ya existentes.

### PREGUNTA 53

El MeSH (*Medical Subject Headings*) es un tesauro que se presenta con una estructura jerárquica.

**Respuesta:** Verdadero. El MeSH se presenta como una lista alfabética y en una estructura jerárquica (v. fig. 12.3), e incluye tanto términos genéricos como específicos, así como cientos de términos relacionados que ayudan a encontrar los descriptores más apropiados. La estructura jerárquica agrupa los descriptores en 15 categorías, en cada una de las cuales se hallan ordenados jerárquicamente, del más general al más específico, hasta siete niveles. Un mismo descriptor puede encontrarse en varias subcategorías o árboles diferentes. Una aplicación interesante de la clasificación jerárquica consiste en que pueden recuperarse, mediante una opción denominada «explosión del término» (explode), no sólo los artículos indizados por dicho descriptor, sino también todos los indizados con los términos más específicos que están subordinados a él.

### PREGUNTA 54

En un estudio analítico, la formulación del objetivo específico requiere como mínimo la identificación de los siguientes elementos: el *factor de estudio*, la *variable de respuesta* o criterio de evaluación, y la *población* o conjunto de sujetos a los que se desea aplicar los resultados del estudio.

**Respuesta:** Verdadero. En un estudio analítico, la formulación del objetivo específico requiere la identificación de los siguientes elementos: el factor de estudio, que corresponde a la exposición o intervención de interés; la variable de respuesta, o criterio de evaluación, con la que se pretende medir el efecto o la asociación con dicho factor, y la población, o conjunto de sujetos a los que se desea aplicar los resultados del estudio. En ocasiones puede ser conveniente incluir en la formulación el factor o los factores que se utilizarán como comparación.

### PREGUNTA 55

Cuando se usa una variable subrogada como variable de respuesta, se hace bajo la asunción de la existencia de una relación directa con la variable de respuesta final.

**Respuesta:** Verdadero. En ocasiones se recurre al uso de una variable intermedia o subrogada en sustitución de la variable primaria de interés porque se cree que está correlacionada con ella, refleja la totalidad o gran parte del efecto, y puede obtenerse con un coste inferior, en un período más breve, o en circunstancias en las que aquella no podría estudiarse. Los cambios observados en la variable intermedia o subrogada se supone que reflejan los

que se producirían en la variable primaria. El uso de estas variables lleva implícita la asunción de la existencia de una relación directa con la respuesta final. Así, el uso de la reducción de las cifras de colesterolemia total tras la administración de un fármaco se hace bajo la asunción de que comporta la disminución de la morbilidad cardiovascular. Si esta asunción no es válida, la variable intermedia o subrogada tampoco lo es, e incluso, si la asunción es válida, estrictamente la conclusión sólo podrá aplicarse a la variable intermedia.

### PREGUNTA 56

La población diana es el subconjunto de la población diana al que se tiene la intención de estudiar. Se define por los criterios de selección establecidos en el protocolo y presenta determinadas características geográficas y temporales que la hacen accesible a los investigadores.

*Respuesta: Falso. Pueden considerarse tres niveles de población:*

1. *Población diana: conjunto de individuos al que hace referencia la pregunta principal u objetivo del estudio. Es la población a la que se desearía generalizar los resultados. Se define principalmente por sus características clínicas y demográficas generales.*
2. *Población de estudio: subconjunto de la población diana al que se tiene la intención de estudiar. Se define por los criterios de selección establecidos en el protocolo y presenta determinadas características geográficas y temporales que la hacen accesible a los investigadores.*
3. *Muestra: conjunto de individuos realmente estudiados. En la mayoría de las ocasiones, el número de sujetos necesarios para la realización del estudio es mucho menor que el de candidatos que forman la población de estudio, por lo que, por razones de eficiencia y disponibilidad de recursos (viabilidad), se selecciona una muestra. En otras ocasiones se incluyen sujetos de forma consecutiva hasta alcanzar el número necesario. En cualquier caso, la estrategia de inclusión de los sujetos debe intentar que la muestra sea representativa de la población de estudio.*

### PREGUNTA 57

Para poder estimar el cálculo del número de individuos necesario para realizar un estudio transversal es imprescindible conocer la variabilidad del parámetro que se desea estimar.

*Respuesta: Verdadero. Es imprescindible conocer la variabilidad del parámetro que se desea estimar. Si no se conoce, puede obtenerse una aproximación a partir de datos propios o de otras investigaciones, o un estudio piloto. En el caso de las variables cuantitativas se mide por la variancia, y en el de las cualitativas, por el producto  $P \cdot (1 - P)$ .*

### PREGUNTA 58

En un estudio transversal, cuanto mayor sea la precisión con que se desea estimar un parámetro, menor es el número de sujetos necesario.

*Respuesta: Falso. Cuanto más precisa se deseé que sea la estimación, más estrecho será el intervalo de confianza, y más sujetos deberán ser estudiados. La precisión debe fijarse previamente, en función de la finalidad de la estimación. En algunos casos puede requerirse una gran precisión, mientras que en otros, si sólo se necesita conocer aproximadamente entre qué valores se encuentra el parámetro, se requerirá una menor precisión y, consecuentemente, menos sujetos.*

### PREGUNTA 59

Si desea estimar un parámetro con una confianza del 99% se necesitarán menos individuos que si se quiere estimar con una confianza del 95%.

*Respuesta: Falso. El nivel de confianza deseado habitualmente se fija en el 95%, correspondiente a un valor  $\alpha$  de 0,05. Este valor indica el grado de confianza que se tiene de que el verdadero valor del parámetro en la población se sitúa en el intervalo obtenido. Cuanta más confianza se deseé, menor será el valor de  $\alpha$ , y más elevado el número de sujetos necesario.*

### PREGUNTA 60

En algunas ocasiones, en un estudio transversal, si no se conoce el valor aproximado del parámetro que se está buscando, puede adoptarse la postura de la máxima indeterminación.

*Respuesta: Verdadero. En algunas ocasiones no se conoce el valor aproximado del parámetro que se está buscando. Si no existen datos de la literatura que resulten útiles, o si no puede realizarse una prueba*

## Métodos de investigación clínica y epidemiológica

piloto para obtener una primera aproximación a dicho valor, puede adoptarse la postura de la máxima indeterminación. Esta actitud consiste en suponer que el porcentaje que se quiere estimar se sitúa alrededor del 50%, ya que es el valor que requiere una mayor cantidad de individuos para una precisión determinada, de forma que el número obtenido supera cualquier otra combinación de parámetros. Esta actitud también es útil, por ejemplo, cuando se realizan encuestas en las que interesa estimar múltiples parámetros. Su inconveniente es que requiere el estudio de un número de sujetos a veces muy superior al realmente necesario.

### PREGUNTA 61

En un estudio en el que se comparan dos o más grupos, un *error tipo I*, conocido también como *error  $\alpha$*  se comete cuando el investigador rechaza la hipótesis nula, siendo ésta verdadera en la población.

**Respuesta:** Verdadero. El *error tipo I*, conocido también como *error  $\alpha$* , se comete cuando el investigador rechaza la hipótesis nula, siendo ésta verdadera en la población. Es equivalente a encontrar un resultado falso positivo, ya que el investigador concluye que hay una diferencia, cuando en realidad no existe.

### PREGUNTA 62

En un muestreo probabilístico, las unidades de muestreo deben coincidir con las unidades sobre las que se realizará el análisis estadístico.

**Respuesta:** Falso. Las unidades de muestreo no tienen que coincidir necesariamente con las unidades de análisis. En un estudio para conocer la prevalencia de HTA en la población escolar, las unidades de muestreo pueden ser las escuelas, pero se analiza en cada una la presión arterial de sus alumnos (unidades de análisis).

### PREGUNTA 63

En un estudio en el que se comparan dos o más grupos, un *error tipo II* se comete cuando el investigador no rechaza la hipótesis nula, siendo ésta falsa en la población.

**Respuesta:** Verdadero. El *error tipo II*, conocido también como *error beta*, es equivalente a un resultado falso negativo, ya que el investigador concluye que ha sido incapaz de encontrar una diferencia o un efecto que existe en la realidad.

### PREGUNTA 64

El proceso de selección en que todos los individuos candidatos tienen una probabilidad conocida, distinta de cero, de ser incluidos en la muestra, se denomina muestreo probabilístico.

**Respuesta:** Verdadero. El muestreo probabilístico se define como el proceso de selección en que todos los individuos candidatos tienen una probabilidad conocida, distinta de cero, de ser incluidos en la muestra. En todas las técnicas probabilísticas, la selección de las unidades se realiza al azar y se evita la posible parcialidad, consciente o inconsciente, de los investigadores. Por esta razón, es más probable que las muestras tiendan a ser representativas, aunque ello no significa que estén exentas de error. En realidad, la muestra seleccionada no es más que una de las muchas que podrían obtenerse de la población de referencia. Es probable que la estimación de la variable obtenida a partir de una muestra difiera de su verdadero valor en la población de referencia por azar (error aleatorio o debido al muestreo), y lo es más cuanto menor es el tamaño de la muestra. El uso de muestras probabilísticas permite la aplicación de técnicas estadísticas capaces de cuantificar este error aleatorio.

### PREGUNTA 65

Si deseamos que la muestra presente la misma distribución que la población en relación con determinadas variables, es preferible usar un muestreo aleatorio estratificado.

**Respuesta:** Verdadero. En este tipo de muestreo, la población se divide en estratos en función de las categorías de las variables por las que se desea estratificar, formando subgrupos con alguna característica en común y mutuamente excluyentes. A continuación, se escoge una muestra aleatoria de cada estrato, manteniendo las proporciones observadas en la población de referencia (muestreo aleatorio estratificado proporcional).

**PREGUNTA 66**

En general, con el muestreo estratificado se obtienen estimaciones menos precisas que con el muestreo aleatorio simple.

**Respuesta:** Falso. En general, con el muestreo estratificado se obtienen estimaciones más precisas que con el muestreo aleatorio simple, porque la variancia total se basa en la de cada uno de los estratos, y éstos son más homogéneos que la población. El aumento de precisión que se consigue depende de las diferencias entre los estratos. Cuanto mayor sea esta diferencia, y menor la variabilidad dentro de cada estrato, mayor será la reducción de la variancia que se produce. Por tanto, al realizar las estimaciones, siempre que se utiliza un muestreo estratificado debe tenerse en cuenta en el análisis.

**PREGUNTA 67**

La validez en el proceso de medición expresa el grado en que el valor que se obtiene se corresponde con el verdadero valor de la variable en un sujeto, es decir, que no existe un sesgo (error sistemático) en el proceso de medición.

**Respuesta:** Verdadero. La validez expresa el grado en que el valor que se obtiene con el proceso de medición se corresponde con el verdadero valor de la variable en un sujeto, es decir, que no existe un sesgo (error sistemático) en el proceso de medición.

**PREGUNTA 68**

La fiabilidad de una medición expresa el grado en que se obtendrían valores similares al aplicar el proceso de medición en más de una ocasión. Es un concepto muy relacionado con el de precisión.

**Respuesta:** Verdadero. La fiabilidad de una medición expresa el grado en que se obtendrían valores similares al aplicar el proceso de medición en más de una ocasión. Se trata de un concepto muy relacionado con el de precisión o variabilidad aleatoria.

**PREGUNTA 69**

Para minimizar la variación debida al observador, una estrategia consiste en repetir la medición varias veces y utilizar alguna medida promedio de los resultados.

**Respuesta:** Falso. La estrategia de repetir la medición varias veces y utilizar alguna medida promedio de los resultados minimiza la variación biológica intraindividual. Otra fuente de variación es el observador. Para reducir la variabilidad debida a los observadores es necesario entrenarlos debidamente antes de iniciar el estudio y elegir medidas lo más objetivas posible, ya que, cuanto más subjetivas sean, mayor es la probabilidad de que exista variabilidad de una medición a otra y entre los distintos observadores. Por último, siempre que sea factible, es conveniente evaluar la variabilidad interobservador e intraobservador.

**PREGUNTA 70**

El error diferencial es aquel que ocurre al medir el factor de estudio y/o la variable de respuesta y que se produce de manera desigual en todos los participantes.

**Respuesta:** Falso. El error diferencial es el que ocurre al medir el factor de estudio y/o la variable de respuesta y que se produce por igual en todos los participantes, con independencia del grupo al que pertenecen o del valor de cualquier otra variable. En otras palabras, significa que cuando la variable es cualitativa, la sensibilidad y la especificidad para medirla permanecen constantes en todos los participantes, al margen de otros factores. Si la variable es continua, lo que permanece constante es la diferencia entre la media observada y la verdadera media.

**PREGUNTA 71**

Cuando se produce un error diferencial en la medición de las variables, el resultado es que se reduce la asociación entre las variables.

**Respuesta:** Falso. El error diferencial ocurre cuando la medición del factor de estudio y/o la variable de respuesta depende del valor de otra variable, y afecta de modo diferente a los participantes en el estudio según el grupo al que pertenecen. Cuando la medida es poco válida, y la sensibilidad y la especificidad de la prueba son diferentes en cada grupo, la mala clasificación que resulta puede sesgar los resultados en cualquier dirección.

## Métodos de investigación clínica y epidemiológica

Una verdadera asociación puede quedar enmascarada, disminuida o aumentada, o bien puede encontrarse un efecto que en realidad no existe.

### PREGUNTA 72

En los ensayos clínicos controlados con placebo donde es posible utilizar las técnicas de enmascaramiento, y más concretamente las de doble ciego, es muy difícil incurrir en un error diferencial.

**Respuesta:** Verdadero. En los ensayos clínicos controlados con placebo donde es posible utilizar las técnicas de enmascaramiento, y más concretamente las de doble ciego, es muy difícil incurrir en un error diferencial, ya que ni los observadores ni los participantes conocen la intervención que están recibiendo. En algunos estudios observacionales es factible mantener ciegos a los observadores, pero no así a los participantes. En esta situación se elimina el error potencial que se deriva de las expectativas de los observadores, pero no el que puede provenir de los participantes. De todos modos, las técnicas de ciego no evitan los errores no diferenciales.

### PREGUNTA 73

La sensibilidad al cambio (responsiveness) será muy importante si un cuestionario se aplica en un estudio sobre opiniones o actitudes acerca de una enfermedad.

**Respuesta:** Falso. La sensibilidad al cambio (responsiveness) será muy importante si el instrumento se aplica como medida de la respuesta en los ensayos clínicos, pero no lo será tanto en un estudio sobre opiniones o actitudes acerca de una enfermedad.

### PREGUNTA 74

La consistencia interna de un cuestionario se refiere a si los ítems que miden un mismo atributo presentan homogeneidad entre ellos.

**Respuesta:** Verdadero. La consistencia interna se refiere a si los ítems que miden un mismo atributo presentan homogeneidad entre ellos. Los cuestionarios se desarrollan para medir separadamente diferentes componentes o dimensiones de un problema. Un cuestionario de salud suele estar dividido en preguntas que tratan de medir la salud física y mental, o un cuestionario de satisfacción en apartados que identifican, por ejemplo, los componentes de competencia profesional, las cualidades personales del profesional sanitario y la accesibilidad de los servicios. En todas estas situaciones es de esperar que exista una buena homogeneidad entre las distintas preguntas que miden un mismo componente. Si en un cuestionario de satisfacción los usuarios contestan que su médico se preocupa bastante de ellos como personas, es de esperar que en otra pregunta sobre si su médico está dispuesto a escucharles contesten afirmativamente. De otro modo se pensará que los distintos ítems que componen la satisfacción con el médico que les atiende son poco consistentes entre sí y que el cuestionario es poco fiable.

### PREGUNTA 75

El alfa de Cronbach es la técnica estadística para estimar la consistencia interna de un cuestionario.

**Respuesta:** Verdadero. La técnica estadística para su análisis es el alfa de Cronbach, que expresa la consistencia interna entre tres o más variables. Sus valores están comprendidos entre 0 y 1, y su interpretación es similar a la de un coeficiente de correlación. Pueden calcularse diferentes valores del alfa de Cronbach excluyendo determinados ítems del cuestionario, de forma que puede evaluarse si la supresión de algunas preguntas mejora la fiabilidad. De todas formas, antes de decidir eliminar un ítem, debe evaluarse si ello puede afectar a la validez del cuestionario, ya que puede ser preferible mantener la pregunta aun a costa de una consistencia interna ligeramente menor.

### PREGUNTA 76

Si el alfa de Cronbach es de 0,10, se puede considerar que el cuestionario tiene una buena consistencia interna.

**Respuesta:** Falso. Como norma general se sugiere que el valor del alfa de Cronbach ha de ser igual o superior a 0,70 para considerar que un instrumento tiene una buena consistencia interna.

**PREGUNTA 77**

Si al analizar la comparabilidad inicial entre dos o más grupos no se encuentran diferencias estadísticamente significativas entre ellos, se puede concluir que los grupos son comparables.

**Respuesta:** Falso. *El análisis de esta comparabilidad no se basa en la significación estadística, ya que algunas diferencias podrían ser estadísticamente significativas sólo por azar, mientras que diferencias relevantes podrían no serlo si el número de sujetos es reducido. Además, los grupos podrían diferir en alguna característica importante, que no hubiera sido medida en el estudio.*

**PREGUNTA 78**

Cuando el efecto o la asociación entre el factor de estudio y la variable de respuesta varían según la existencia de una tercera variable o de diferentes niveles de ella, se dice que ésta es confusora de dicho efecto.

**Respuesta:** Falso. *Cuando el efecto o la asociación entre el factor de estudio y la variable de respuesta varían según la existencia de una tercera variable o de diferentes niveles de ella, se dice que ésta es modificadora de dicho efecto. Estas variables son importantes porque forman parte del propio efecto del factor de estudio, por lo que, más que controlar por ellas, interesa tenerlas en cuenta en el análisis para poder describir cómo modifican dicho efecto.*

**PREGUNTA 79**

Si en el análisis estadístico de un estudio se detecta la existencia de modificación del efecto, ésta debería controlarse.

**Respuesta:** Falso. *La estrategia de análisis ha de planificarse de forma que permita obtener la mejor estimación posible de la magnitud del efecto del factor de estudio sobre la variable de respuesta, teniendo en cuenta las influencias del resto de variables que pueden interferir. En primer lugar, deben identificarse las variables que pueden ser modificadoras del efecto, prever la estimación del efecto en los diferentes estratos definidos por los valores de dichas variables, y la aplicación de pruebas estadísticas para comparar las estimaciones obtenidas. Si se detecta la existencia de modificación del efecto, deberá tenerse en cuenta al presentar los resultados, ya que no existirá un único resultado, sino que variará en función de las categorías o los valores de otras variables. Si no se detecta ninguna modificación del efecto, deberá obtenerse una estimación del resultado ajustada por los potenciales factores de confusión.*

**PREGUNTA 80**

Si las no respuestas se producen de forma aleatoria o se deben a motivos que no están relacionados con el problema en estudio, la consecuencia será que la estimación que se obtenga tendrá una menor precisión, o la comparación una menor potencia estadística, ya que se habrá estudiado un número menor de individuos.

**Respuesta:** Verdadero. *Conceptualmente, el problema de las no respuestas radica no tanto en su número o en el porcentaje que suponen respecto a la muestra, sino en los motivos por los que se han producido. Si están asociadas al problema en estudio o a determinadas características relacionadas con éste, el conjunto de sujetos analizados no es representativo de la población que se deseaba estudiar, y el resultado obtenido no será una buena estimación del verdadero valor de dicha población. Si, por el contrario, las no respuestas se producen de forma aleatoria o se deben a motivos que no están relacionados con el problema en estudio, la consecuencia será que la estimación que se obtenga tendrá una menor precisión, o la comparación una menor potencia estadística, ya que se habrá estudiado un número menor de individuos, pero no estará sesgada.*

**PREGUNTA 81**

Con el fin de preservar la comparabilidad de los grupos, la opción más válida de análisis estadístico es evaluar a todos los pacientes incluidos en el estudio según el principio denominado *análisis según intención de tratar*.

**Respuesta:** Verdadero. *Dado que las pérdidas y retiradas difícilmente se producen al azar, cualquier exclusión de sujetos del análisis puede alterar la comparabilidad de los grupos. Con el fin de preservarla, la única opción válida de análisis es evaluar a todos los pacientes incluidos en el estudio según el principio denominado análisis*

## Métodos de investigación clínica y epidemiológica

según intención de tratar (intention-to-treat analysis) o según asignación aleatoria (as-randomized), según el cual se analiza a todos los pacientes como pertenecientes al grupo al que fueron asignados inicialmente, con independencia del tratamiento que hayan recibido realmente o de cualquier desviación del protocolo que se haya producido. Algunos autores matizan la definición de este principio, incluyendo en el análisis tan sólo a los sujetos que han recibido al menos una dosis del tratamiento. Aunque pueda parecer extraño analizar en un grupo a sujetos que no han tomado el tratamiento asignado, o incluso que han recibido la intervención del otro grupo u otras medicaciones, si no se siguiese el principio de intención de tratar, se comprometería la comparabilidad de los grupos. De todas formas, si estas situaciones se producen en un gran número de sujetos, el estudio debería ser considerado no evaluable.

### PREGUNTA 82

Un análisis estadístico exploratorio sirve para contrastar las hipótesis planteadas en un estudio.

**Respuesta:** Falso. Las técnicas de análisis estadístico pueden utilizarse para explorar conjuntos de datos sin hipótesis previas o bien para confirmar hipótesis de trabajo. Ambas finalidades están vinculadas a la naturaleza de los objetivos del estudio, a la actitud con que el investigador se enfrenta a los datos y a los términos en que deberán interpretarse los resultados. Una hipótesis solamente puede confirmarse mediante un estudio diseñado con el propósito de hacerlo. La exploración implica el rastreo de datos en busca de información, sin objetivos concretos ni hipótesis que hayan gobernado el diseño del estudio. La exploración puede servir para sugerir nuevas hipótesis, pero de ningún modo para contrastarlas, sino que la confirmación deberá obtenerse en un nuevo estudio diseñado para ello.

### PREGUNTA 83

Cuando en un estudio se concluye que existe una diferencia cuando en realidad no la hay, se comete un error tipo II o error  $\beta$ .

**Respuesta:** Falso. En un estudio, puede concluirse que existe una diferencia cuando en realidad no la hay. Es decir, puede rechazarse la  $H_0$  cuando en realidad es cierta. Si esto ocurre, la decisión es incorrecta y se comete un error, conocido como error tipo I o error  $\alpha$ . La probabilidad de cometer este error es la de que, si se concluye que existe una diferencia significativa, ésta sea en realidad debida al azar. Si se hace un símil entre una prueba estadística y una diagnóstica, equivale a la probabilidad de obtener un resultado falso positivo. Esto es precisamente lo que mide el valor de  $p$ , o grado de significación estadística de la prueba. Si, por el contrario, se concluye que la diferencia no es estadísticamente significativa, es decir, si no se rechaza la  $H_0$ , puede ocurrir que la hipótesis sea falsa y que, en realidad, exista una diferencia entre ambos grupos, en cuyo caso se cometrá otro tipo de error, llamado error tipo II o  $\beta$ .

### PREGUNTA 84

Un resultado estadísticamente significativo no implica que sea clínicamente relevante.

**Respuesta:** Verdadero. Un resultado estadísticamente significativo no implica que sea clínicamente relevante. El valor de  $p$  no mide la fuerza de la asociación. Pueden obtenerse valores pequeños de  $p$  (resultados estadísticamente significativos) simplemente estudiando un número grande de sujetos. Al aumentar el tamaño de la muestra se incrementa el poder estadístico para detectar incluso pequeñas diferencias. La diferencia que se considera clínicamente relevante depende de su magnitud y de otros factores, como la frecuencia y la gravedad de los efectos secundarios de ambos fármacos, la facilidad de administración o su coste económico.

### PREGUNTA 85

Cuando se utiliza como medida del efecto una diferencia, si el IC del 95% incluye el valor 1, que es el que corresponde a la  $H_0$ , se concluirá que el resultado no es estadísticamente significativo.

**Respuesta:** Falso. Cuando se utiliza como medida del efecto una diferencia, si el IC del 95% incluye el valor 0, que es el que corresponde a la  $H_0$  (ausencia de diferencia entre ambos grupos), se concluirá que el resultado no es estadísticamente significativo, ya que no puede descartarse que ése sea el verdadero valor. Si, por el contrario, el IC excluye este valor 0, se concluirá que la diferencia observada es estadísticamente significativa, ya que puede descartarse dicho valor (con un riesgo de error inferior al 5%). Por tanto, el IC, además de indicar si la diferencia es o no estadísticamente significativa, permite conocer entre qué límites es probable que se encuentre la verdadera diferencia, lo que es muy útil en la interpretación de los resultados.

**PREGUNTA 86**

La potencia estadística cuantifica la capacidad de un estudio para detectar como estadísticamente significativa una determinada diferencia o asociación que existe en la realidad.

**Respuesta:** Verdadero. La probabilidad de no detectar como estadísticamente significativa una diferencia o asociación que existe en la realidad se conoce como error  $\beta$ . Su complementario  $1 - \beta$  corresponde a la potencia estadística, que cuantifica, por tanto, la capacidad de un estudio para detectar como estadísticamente significativa una determinada diferencia o asociación que existe en la realidad.

**PREGUNTA 87**

El uso de variables combinadas para medir un efecto permite aumentar la potencia estadística.

**Respuesta:** Verdadero. El uso de variables combinadas para medir un efecto permite aumentar la potencia estadística, aunque debe tenerse en cuenta que su utilización puede diluir determinadas diferencias o asociaciones que afecten a algunos de sus componentes.

**PREGUNTA 88**

El sesgo de selección debido a la supervivencia selectiva se produce en los estudios en que se incluyen sujetos que han sobrevivido a las primeras fases de su enfermedad.

**Respuesta:** Verdadero. El sesgo de selección debido a la supervivencia selectiva se produce en los estudios en que se incluyen casos prevalentes, es decir, sujetos que han sobrevivido a las primeras fases de su enfermedad. Si la letalidad de una enfermedad depende de la exposición (p. ej., si la cardiopatía isquémica es más grave en los sujetos que presentan más factores de riesgo), los casos prevalentes no serán representativos de todos los casos, sino que presentarán menores niveles de exposición, lo que puede conducir a una estimación sesgada de la asociación. Además, los casos prevalentes pueden haber cambiado sus hábitos en relación con la exposición (p. ej., haber dejado de fumar tras un infarto de miocardio), hecho que también conduciría a un sesgo en los resultados.

**PREGUNTA 89**

El sesgo de memoria es un error de clasificación no diferencial.

**Respuesta:** Falso. Un ejemplo de error de clasificación diferencial es el denominado sesgo de memoria, que se produce cuando los individuos con un determinado problema de salud recuerdan su historia de exposición de forma distinta a como la recuerdan aquellos que no tienen dicho problema. Suele producirse cuando se recoge la información sobre la exposición mediante una entrevista o un cuestionario, y los casos recuerdan sus antecedentes de forma diferente a los controles. También puede producirse cuando los individuos expuestos a algún riesgo (ocupacional, ambiental, etc.) informan sobre el desarrollo de problemas de salud de forma diferente a como lo hacen los que no están expuestos. Este tipo de sesgo es especialmente problemático en los estudios de casos y controles, y en los retrospectivos de cohortes, ya que tanto la exposición como la enfermedad ya han ocurrido, y puede conducir tanto a una sobreestimación como a una infraestimación de la asociación entre la exposición y la enfermedad, dependiendo de si los casos recuerdan su exposición mejor o peor que los controles.

**PREGUNTA 90**

Si un encuestador conoce la hipótesis puede incurrir en un error de clasificación diferencial.

**Respuesta:** Verdadero. Un encuestador puede introducir un error de clasificación diferencial si conoce la hipótesis del estudio y la condición (de expuesto o enfermo, según el tipo de estudio) del entrevistado. En esta situación puede obtener información de los individuos de forma diferente (más exacta o más exhaustiva) según el grupo de estudio al que pertenezcan.

**PREGUNTA 91**

La identificación de un factor de confusión debe basarse en la existencia de una diferencia estadísticamente significativa entre los grupos de estudio.

**Respuesta:** Falso. La confusión puede definirse como la distorsión que se produce en la estimación de la magnitud de la asociación entre el factor de estudio y la variable de respuesta, debida a la influencia de uno o varios factores externos. Para producir confusión, una variable debe estar asociada tanto con el factor de estudio como con la variable de respuesta, y no debe formar parte de la cadena causal entre ellos. Por tanto, deben considerarse

## Métodos de investigación clínica y epidemiológica

potenciales factores de confusión las variables por las que difieren los grupos que se comparan y que son pronósticas de la respuesta. Su identificación no debe basarse en la existencia de una diferencia estadísticamente significativa entre los grupos de estudio. Dado que la significación estadística depende tanto de la magnitud de la diferencia observada como del tamaño de la muestra, una diferencia muy pequeña e irrelevante puede ser estadísticamente significativa si el tamaño de los grupos es muy grande, y una diferencia de gran importancia puede ser estadísticamente no significativa si el número de sujetos incluidos es muy pequeño. Por lo tanto, la significación estadística no es un buen criterio para considerar que una variable puede ser un potencial factor de confusión. Lo que cuenta es la posible relevancia que pueda tener sobre los resultados la diferente distribución de una variable entre los grupos. Además, no debe olvidarse que múltiples diferencias de escasa importancia aparente pueden producir una distorsión clínicamente relevante del resultado al actuar de forma conjunta.

### PREGUNTA 92

La presencia de un fenómeno de confusión puede detectarse comprobando si el análisis bruto (sin tener en cuenta el factor de confusión) y el ajustado (controlando el potencial fenómeno de confusión) conducen a resultados que difieren de forma clínicamente relevante.

**Respuesta:** Verdadero. Desde un punto de vista práctico, la presencia de un fenómeno de confusión puede detectarse comprobando si el análisis bruto (sin tener en cuenta el factor de confusión) y el ajustado (controlando el potencial fenómeno de confusión) conducen a resultados que difieren de forma clínicamente relevante. La evaluación de la confusión no se basa, pues, en una prueba estadística, sino en la valoración de lo relevante que pueda considerarse la diferencia entre ambas estimaciones. Aunque es difícil fijar un criterio único para concluir que ha existido confusión, ya que depende de cada situación concreta, algunos autores lo establecen en un cambio de un 15-20% en la estimación del efecto de interés. Sin embargo, debe tenerse en cuenta que lo importante no es tanto identificar si ha existido confusión, sino, en caso de que se haya producido, obtener una estimación no distorsionada de la magnitud del efecto del factor de estudio sobre la variable de respuesta, es decir, que se haya controlado su influencia (estimación ajustada).

### PREGUNTA 93

El ajuste por variables que no son verdaderos factores de confusión sesga la estimación, y la hace más imprecisa.

**Respuesta:** Falso. El ajuste por variables que no son verdaderos factores de confusión no sesga la estimación, pero la hace más imprecisa (fenómeno que se conoce como sobreajuste).

### PREGUNTA 94

Una variable modificadora del efecto debe ser descrita, no controlada. Una variable confusora debe ser descrita y controlada.

**Respuesta:** Verdadero. Los fenómenos de confusión y de modificación del efecto son muy distintos tanto en la información que proporcionan como en la forma de interpretarla. Si una variable actúa como un factor de confusión depende de cómo se distribuye entre los grupos de estudio y de si el efecto observado del factor de estudio sobre la variable de respuesta puede explicarse total o parcialmente por esta distribución desigual. Por lo tanto, una variable puede actuar como factor de confusión en un estudio experimental, pero no en otro, dependiendo de cómo quede distribuida tras el proceso de asignación. En cambio, en los estudios observacionales, al estar asociados a la exposición y a la variable de respuesta, los factores de confusión siempre actúan como tales cuando no se ha utilizado una técnica para prevenir su efecto (restricción o emparejamiento). Cuando una variable se comporta como un factor de confusión, el interés reside en controlar su efecto. En cambio, la modificación del efecto refleja una característica de la relación entre el factor de estudio y la variable de respuesta, ya que la magnitud de esta relación depende de una tercera variable. Por tanto, la modificación del efecto debe ser descrita, pero no controlada. Además, puede proporcionar claves interesantes (hipótesis de trabajo) sobre los mecanismos causales de dicha relación.

### PREGUNTA 95

La fuerza de una asociación entre dos variables no se mide por el grado de significación estadística.

**Respuesta:** Verdadero. La fuerza de una asociación no se mide por el grado de significación estadística. El valor de  $p$  depende de la magnitud de la asociación y del tamaño de la muestra, de forma que asociaciones de escasa o nula relevancia pueden aparecer como estadísticamente significativas en estudios de gran tamaño.

Por otro lado, esta valoración no debe hacerse solamente según la estimación puntual de la magnitud de la asociación, sino que debe tenerse en cuenta la precisión con que se ha valorado, es decir, su intervalo de confianza. Es frecuente encontrar estudios realizados con un número reducido de sujetos, que presentan riesgos relativos cuyo intervalo de confianza tiene un límite superior muy elevado. Hay que interpretar con cautela estos resultados, ya que reflejan más una ausencia de precisión que una asociación muy fuerte.

### PREGUNTA 96

El hecho de que la magnitud de la respuesta aumente a medida que se incrementa el grado de exposición al factor de estudio (en cantidad y/o tiempo) es un criterio imprescindible para concluir que la relación entre ambos es causal.

**Respuesta:** Falso. El hecho de que la magnitud de la respuesta aumente a medida que se incrementa el grado de exposición al factor de estudio (en cantidad y/o tiempo) es un argumento más a favor de que la relación entre ambos es causal. Para evaluar este criterio, la exposición al factor de estudio debe tener más de dos categorías, y cada una de ellas debe reflejar los incrementos de la exposición; se trata de determinar si implican también aumentos del riesgo. Algunos autores incluyen este criterio como un elemento de la plausibilidad biológica. Sin embargo, no todas las relaciones causales deben tener necesariamente un gradiente dosis-respuesta. Puede ocurrir, por ejemplo, que exista un umbral por encima del cual aumente el riesgo, sin hacerlo más aunque aumente el grado de exposición. Así mismo, la relación no tiene por qué ser lineal; se puede adoptar otro tipo de relación que no se ponga de manifiesto con las técnicas de análisis habituales.

### PREGUNTA 97

La diferencia de incidencias (DI) o exceso de riesgo se define como la diferencia entre la incidencia en los expuestos y la incidencia en los no expuestos al factor de estudio. La DI representa la cantidad de incidencia que puede ser atribuida al factor de estudio.

**Respuesta:** Verdadero. La diferencia de incidencias (DI) o exceso de riesgo se define como la diferencia entre la incidencia en los expuestos (I1) y la incidencia en los no expuestos al factor de estudio (I0) (v. tabla 32.1). La DI representa la cantidad de incidencia que puede ser atribuida al factor de estudio.

### PREGUNTA 98

El riesgo relativo (RR) no tiene en cuenta la magnitud de las incidencias que se están comparando, mientras que la diferencia de incidencias (DI) sí tiene en cuenta la magnitud de las incidencias que se están comparando.

**Respuesta:** Verdadero. La DI representa la cantidad de incidencia que puede ser atribuida al factor de estudio. A diferencia de la DI, el RR no tiene en cuenta la magnitud de las incidencias que se están comparando.

### PREGUNTA 99

El riesgo relativo (RR) da una idea del impacto que tendría sobre la incidencia de una enfermedad la eliminación de un factor de riesgo.

**Respuesta:** Falso. La diferencia de incidencias (DI) y el RR son dos formas diferentes (y complementarias) de abordar un mismo problema. Bajo el punto de vista de la salud pública, la DI es más informativa, ya que muestra el exceso de casos que están asociados a la exposición al factor de estudio. Por otro lado, desde el punto de vista de la investigación etiológica, el RR es más informativo, ya que estima la fuerza de la asociación entre el factor de estudio y la respuesta. La DI no sólo depende del RR, sino también de la frecuencia de la enfermedad. Para enfermedades poco frecuentes, aun cuando el RR asociado a una exposición sea elevado, la DI puede ser más pequeña que si el RR de esa misma exposición es menor, pero la enfermedad mucho más frecuente.

### PREGUNTA 100

El número de pacientes que es necesario tratar (NNT) es una medida del impacto.

**Respuesta:** Verdadero. Una medida que se puede obtener en los ensayos clínicos, fácil de calcular a partir de la disminución absoluta de riesgo, comprensible para el clínico y extrapolable a la realidad de su labor diaria,

## **Métodos de investigación clínica y epidemiológica**

*es el número de casos que es necesario tratar (NNT) para prevenir un episodio. Desde el punto de vista matemático, esta medida equivale al recíproco del valor de la reducción absoluta de riesgo y, por tanto, presenta las mismas ventajas en la expresión de resultados. Además, presenta la ventaja adicional de que indica el esfuerzo (impacto) terapéutico que debe realizarse para poder prevenir un episodio.*

# Glosario

<b>Abandono</b>	<i>Drop-out</i>	Sujeto que no completa el seguimiento de un estudio, ya sea voluntariamente o por motivos no relacionados con el propio estudio. Ver también <i>pérdidas de seguimiento</i> .
<b>Acontecimiento adverso</b>	<i>Adverse event</i>	Cualquier experiencia no deseable que ocurra a un sujeto durante un <i>ensayo clínico</i> , se considere o no relacionada con los productos en investigación. Cuando el fármaco ya está comercializado, se prefiere el término <i>reacción adversa</i> .
<b>Adaptación transcultural</b>	<i>Transcultural adaptation</i>	Adaptación de un <i>cuestionario</i> a otra cultura con el objetivo de conseguir un instrumento equivalente al desarrollado en el país de origen. El método que suele utilizarse es el de la traducción-retraducción por personas bilingües, seguido de un análisis de la nueva versión para detectar discrepancias y de la comprobación de su comprensión y aceptabilidad por parte de un grupo de sujetos.
<b>Adherencia</b>	<i>Adherence, compliance</i>	Ver <i>cumplimiento</i> .
<b>Ajuste</b>	<i>Adjustment</i>	Procedimiento estadístico por el que se obtiene una estimación de un efecto o asociación en la que se ha controlado (eliminado) la influencia de los <i>factores de confusión</i> . Puede realizarse mediante <i>análisis estratificado</i> , <i>estandarización</i> o <i>análisis multivariante</i> .
<b>Aleatorio</b>	<i>Random</i>	Regido por el azar. Ver también <i>asignación aleatoria</i> y <i>muestreo aleatorio</i> .
<b>Aleatorización</b>	<i>Randomization</i>	Ver <i>asignación aleatoria</i> .
<b>Alfa de Cronbach</b>	<i>Cronbach's alpha</i>	Procedimiento estadístico que se utiliza para evaluar la <i>consistencia interna</i> de un cuestionario. Sus valores oscilan entre 0 y 1, y su interpretación es similar a la de un coeficiente de correlación.
<b>Algoritmo</b>	<i>Algorithm</i>	Proceso sistemático que consiste en una secuencia ordenada de pasos, cada uno de los cuales depende del resultado del paso previo.
<b>Ambispectivo</b>	<i>Ambispective</i>	Diseño que combina una parte <i>retrospectiva</i> con otra <i>prospectiva</i> .
<b>Amplitud</b>	<i>Range</i>	De una serie de valores, ver <i>recorrido</i> .

## Glosario

<b>Análisis coste-beneficio</b>	<i>Cost-benefit analysis</i>	Evaluación de los efectos de una intervención, convertidos en los mismos términos monetarios que los costes de la misma, con los que los compara.
<b>Análisis coste-efectividad</b>	<i>Cost-effectiveness analysis</i>	Evaluación de los efectos de una intervención, expresados en términos sanitarios, y que describe los costes por una ganancia adicional (p. ej., por infarto de miocardio preventido o por año de vida ganado).
<b>Análisis coste-utilidad</b>	<i>Cost-utility analysis</i>	Evaluación de los efectos de una intervención, convertidos en preferencias personales o <i>utilidades</i> , y que describe los costes por una ganancia adicional (p. ej., por año de vida ajustado por calidad o AVAC adicional).
<b>Análisis de clusters</b>	<i>Cluster analysis</i>	Técnica de <i>análisis multivariante</i> que tiene por finalidad determinar si un conjunto de sujetos pueden ser clasificados en grupos, de forma que los sujetos de cada uno de ellos se parezcan más entre sí en los valores que toman para las variables de análisis que a los sujetos de otros grupos. No existe, pues, una variable dependiente. Ver también <i>análisis de correspondencias</i> .
<b>Análisis de componentes principales</b>	<i>Principal component analysis</i>	Técnica de <i>análisis multivariante</i> que tiene por finalidad reducir un conjunto de variables interrelacionadas en un conjunto menor de variables no correlacionadas, llamadas componentes principales. No existe una <i>variable dependiente</i> . A menudo se utiliza como un primer paso de un <i>análisis factorial</i> .
<b>Análisis de correspondencias</b>	<i>Correspondence analysis</i>	Técnica de <i>análisis de clusters</i> en que las variables son categóricas.
<b>Análisis de decisión</b>	<i>Decision analysis</i>	Aplicación de métodos cuantitativos explícitos para analizar las decisiones en situaciones de incertidumbre, a través de una representación y un examen sistemáticos de toda la información y toda la incertidumbre relacionadas con aquella información. Las opciones disponibles se representan en un árbol de decisiones. En cada rama o nodo de decisiones, las probabilidades de cada resultado que se puede predecir quedan estimadas.
<b>Análisis de la covarianza</b>	<i>Analysis of covariance, ANCOVA</i>	Prueba estadística para la comparación de las <i>medias</i> de una variable en diferentes poblaciones, que incorpora el <i>ajuste</i> por una o varias <i>variables concomitantes</i> , asumiendo que no existe <i>interacción</i> con el factor de estudio.
<b>Análisis de la varianza</b>	<i>Analysis of variance, ANOVA</i>	1. De una vía: prueba estadística para la comparación de más de dos <i>medias</i> en <i>muestras independientes</i> , cuando las variables siguen una <i>distribución normal</i> y las <i>varianzas</i> son iguales ( <i>homoscedasticidad</i> ). Su base es que la variabilidad total de los datos tiene dos componentes: la variabilidad dentro de cada muestra (debida al azar) y la variabilidad existente entre las muestras (debida al azar y a las diferencias entre las muestras). La prueba compara ambos componentes. Si la <i>hipótesis nula</i> es cierta, y no hay diferencias entre las muestras, ambos componentes coincidirán. Si se observa un resultado estadísticamente significativo, indica que hay diferencias entre las muestras en global.

<b>Análisis de la varianza multivariante</b>	<i>Multivariate analysis of variance, MANOVA</i>	Para identificar qué muestra o muestras difieren, deben utilizarse procedimientos estadísticos específicos para realizar <i>comparaciones múltiples</i> (ver <i>Scheffé, Tukey o Newman-Keuls</i> ). 2. De dos vías: prueba estadística para la comparación de más de dos medias en muestras apareadas, cuando las variables siguen una distribución normal y las varianzas son iguales (homoscedasticidad). Su base es que la variabilidad total de los datos tiene tres componentes: la variabilidad existente entre los grupos, la existente entre los sujetos y la debida al azar, denominada error o residual. La prueba compara los dos primeros componentes con el residual.
<b>Análisis de minimización de costes</b>	<i>Cost-minimization analysis</i>	Evaluación económica de diferentes intervenciones que se supone que tienen los mismos efectos, por lo que se comparan únicamente sus costes.
<b>Análisis de sensibilidad</b>	<i>Sensitivity analysis</i>	Procedimiento en que los resultados de un estudio son recalcados según diferentes asunciones o valores de determinadas variables, con la finalidad de ver si alteran las conclusiones del mismo. Se utiliza para valorar cuán robustos son los resultados en cuanto a las decisiones inciertas o las asunciones realizadas con los datos y los métodos que se utilizaron.
<b>Análisis de subgrupos</b>	<i>Subgroup analysis</i>	Análisis realizado en subpoblaciones de los sujetos participantes en un estudio, definidos por determinadas variables. Sólo está justificado si se ha planeado antes del inicio del estudio, en función de una hipótesis fundamentada. Los resultados de este análisis deben interpretarse con cautela, teniendo en cuenta los problemas derivados de las <i>comparaciones múltiples</i> .
<b>Análisis de supervivencia</b>	<i>Survival analysis</i>	Procedimientos estadísticos de análisis que tienen en cuenta el número de acontecimientos observados (muerte u otros) en un grupo de sujetos y el momento en el tiempo en que ocurren en relación con el momento del inicio del seguimiento de cada sujeto.
<b>Análisis discriminante</b>	<i>Discriminant analysis</i>	Técnica de <i>análisis multivariante</i> que se utiliza en el caso de una <i>variable dependiente categórica</i> . Dado que realiza asunciones estrictas sobre la distribución de las variables, cada vez se utiliza más la <i>regresión logística</i> en esta situación.
<b>Análisis económico</b>	<i>Economic analysis</i>	Comparación de los costes y resultados de las intervenciones sanitarias alternativas. Ver <i>análisis coste-beneficio, análisis coste-efectividad, análisis coste-utilidad</i> y <i>análisis de minimización de costes</i> .
<b>Análisis estratificado</b>	<i>Stratified analysis</i>	Evaluación del efecto o asociación entre el <i>factor de estudio</i> y la <i>variable de respuesta</i> en diferentes categorías homogéneas o estratos, definidas según los valores de una o más variables. Cuando la magnitud del efecto es diferente entre los estratos, se dice que ha existido <i>modificación de efecto</i> .

<b>Análisis factorial</b>	<i>Factor analysis</i>	Cuando es similar entre ellos, puede calcularse una medida sintética del efecto, ajustada por la o las variables por las que se ha estratificado. Si esta medida sintética difiere de la obtenida con los datos crudos, las variables de estratificación producirán un fenómeno de <i>confusión</i> , por lo que la medida ajustada es la mejor estimación de la medida del efecto.
<b>Análisis multivariante</b>	<i>Multivariate analysis</i>	Técnica de <i>análisis multivariante</i> que se utiliza para analizar las relaciones entre un conjunto de variables con la finalidad de determinar el número de dimensiones fundamentales que subyacen en los datos observados. Se utiliza con frecuencia en el desarrollo de sistemas de puntuación y ponderación de escalas de medida y cuestionarios. No existe una <i>variable dependiente</i> .
<b>Análisis intermedio</b>	<i>Interim analysis</i>	Procedimientos estadísticos utilizados para estudiar simultáneamente las relaciones entre diferentes variables.
<b>Análisis por intención de tratar</b>	<i>Intention-to-treat analysis (ITT analysis)</i>	Análisis realizado durante la realización de un estudio con la finalidad de monitorizar los resultados. Su realización debe establecerse previamente en el diseño del estudio para evitar los problemas derivados de las <i>comparaciones múltiples</i> .
<b>Análisis según protocolo</b>	<i>Per-protocol analysis</i>	Estrategia de análisis de los resultados de un <i>ensayo clínico aleatorio</i> en que cada sujeto se mantiene en el grupo que le corresponde según la <i>asignación aleatoria</i> , con independencia de si ha cumplido con la intervención asignada o ha recibido otro tratamiento, o de si el seguimiento ha sido completo. Según algunos autores, deben incluirse solamente los sujetos que han recibido como mínimo una dosis de la intervención y han realizado como mínimo una visita de seguimiento. Opuesto a <i>análisis según protocolo</i> .
<b>Analítico</b>	<i>Analytic</i>	Estrategia de análisis de los resultados de un <i>ensayo clínico aleatorio</i> en que sólo se analizan los sujetos que han recibido la intervención asignada, han completado el seguimiento y han cumplido el protocolo de estudio (casos válidos). Opuesto a <i>análisis por intención de tratar</i> .
<b>Analogía</b>	<i>Analogy</i>	Ver <i>estudio analítico</i> .
<b>ANCOVA</b>	<i>ANCOVA</i>	Criterio de <i>causalidad</i> que se refiere a la semejanza de la relación observada con otra cuya causalidad está bien establecida.
<b>ANOVA</b>	<i>ANOVA</i>	Ver <i>análisis de la covarianza</i> .
<b>Años de vida ajustados por calidad (AVAC)</b>	<i>Quality adjusted life years (QALY)</i>	Esperanza de vida ponderada según la calidad de vida y las preferencias personales o <i>utilidades</i> .
<b>Apareamiento</b>	<i>Matching</i>	Ver <i>emparejamiento</i> .
<b>Aplicabilidad</b>	<i>Applicability</i>	Grado en que los resultados de un estudio son válidos en otros ámbitos. Ver <i>valididad externa</i> .

<b>Apuntamiento</b>	<i>Kurtosis</i>	Ver <i>kurtosis</i> .
<b>Área bajo la curva</b>	<i>Area under the curve (AUC)</i>	1. Método para sintetizar la información de una serie de medidas de un sujeto, muy utilizado en farmacología clínica.
<b>Asignación</b>	<i>Assignment</i>	Distribución de los participantes en un estudio en cada uno de los grupos.
<b>Asignación adaptativa</b>	<i>Adaptive assignment</i>	En un <i>ensayo clínico</i> , técnicas de <i>asignación</i> de los sujetos a los grupos de estudio en las que la probabilidad de asignación se va modificando, ya sea en función de las diferencias en la distribución de las variables pronósticas o en el número de sujetos de cada uno de los grupos, o de la respuesta observada en los pacientes previamente estudiados, intentando minimizar el número de sujetos que reciben la intervención menos eficaz.
<b>Asignación aleatoria</b>	<i>Randomization</i>	Técnica que consiste en distribuir los participantes en un ensayo clínico a los grupos de estudio siguiendo un método regido por el azar, lo que tiende a conseguir que las variables que pueden influir en el resultado, ya sean conocidas o no, se distribuyan de forma equilibrada en los diferentes grupos de estudio, especialmente cuando el número de sujetos es elevado. Ver también <i>ensayo clínico aleatorio</i> .
<b>Asignación estratificada</b>	<i>Stratified assignment</i>	En un <i>ensayo clínico aleatorio</i> , técnica de <i>asignación</i> de los sujetos a los grupos de estudio basada en la división de la muestra en subgrupos en función de las variables pronósticas consideradas, procediéndose a continuación a la <i>asignación aleatoria</i> de los sujetos dentro de cada uno de los estratos. Esta técnica ayuda a prevenir la aparición de desequilibrios en la distribución de las variables pronósticas.
<b>Asignación por bloques</b>	<i>Blocks assignment</i>	En un <i>ensayo clínico</i> , técnica de <i>asignación</i> de los sujetos a los grupos de estudio, basada en la creación de unos bloques de un tamaño prefijado, múltiplo de dos, de forma que la mitad de los sujetos de cada uno de estos bloques será asignada a un grupo y la otra mitad al otro. El orden en que estas intervenciones son asignadas dentro de los bloques es aleatorio. Esta técnica garantiza que, en cualquier momento a lo largo del proceso, el número de sujetos de cada uno de los grupos será muy similar (la máxima diferencia será la mitad del tamaño de los bloques).
<b>Asignación sistemática</b>	<i>Systematic assignment</i>	En un <i>ensayo clínico</i> , <i>asignación</i> de los sujetos a los grupos de estudio de forma alternativa según la secuencia ABABA-BAB, la fecha de nacimiento, número de historia clínica par o impar, etc. Este método no es aleatorio, y no es aconsejable su uso, ya que el investigador conoce qué tratamiento recibirá el próximo sujeto que entre en el estudio, lo que puede influir en su decisión de incluirlo o no.

## Glosario

<b>Asimetría</b>	<i>Skewness</i>	Grado en que las colas de una <i>distribución unimodal</i> no son simétricas. Cuando la cola derecha es más larga, se habla de asimetría positiva, y cuando lo es la izquierda, de asimetría negativa. Algunos autores utilizan el término <i>sesgo</i> como sinónimo de asimetría.
<b>Asociación</b>	<i>Association</i>	Grado en que dos variables se relacionan con más frecuencia de la que cabe esperar por azar. Existe asociación cuando la frecuencia de un suceso, la presencia de una característica o el valor de una variable dependen de la frecuencia de uno o más sucesos, la presencia de una o más características o el valor de una o más variables. La asociación entre dos variables es positiva cuando los valores de una de ellas aumentan a medida que lo hacen los de la otra, y negativa cuando aumentan a medida que disminuyen los valores de la otra variable. Opuesto a <i>independencia</i> .
<b>Auditoría</b>	<i>Audit</i>	Examen o revisión que trata de determinar el grado en que una condición o proceso se adecúa a estándares o criterios pre establecidos.
<b>Aumento absoluto de beneficio (AAB)</b>	<i>Absolute Benefit Increase (ABI)</i>	<i>Medida del efecto</i> que se utiliza en los <i>ensayos clínicos aleatorios</i> cuando la intervención experimental aumenta la probabilidad de una respuesta positiva o un acontecimiento deseable. Corresponde a la diferencia entre las incidencias de dicho efecto positivo observadas en el grupo de intervención y en el grupo control ( $AAB = I_e - I_0$ ). Ver <i>diferencia de incidencias</i> .
<b>Aumento absoluto de riesgo (AAR)</b>	<i>Absolute Risk Increase (ARI)</i>	<i>Medida del efecto</i> que se utiliza en los <i>ensayos clínicos aleatorios</i> cuando la intervención experimental aumenta el riesgo de un acontecimiento adverso o indeseable. Corresponde a la diferencia entre las incidencias de dicho efecto negativo observadas en el grupo de intervención y en el grupo control ( $AAR = I_e - I_0$ ). Ver <i>diferencia de incidencias</i> .
<b>Aumento relativo de beneficio (ARB)</b>	<i>Relative Benefit Increase (RBI)</i>	<i>Medida del efecto</i> que se utiliza en los <i>ensayos clínicos aleatorios</i> cuando la intervención experimental aumenta la probabilidad de una respuesta positiva o de un acontecimiento deseable. Expresa el aumento del beneficio obtenido en relación con el grupo control. Corresponde al cociente entre el aumento absoluto del beneficio y la incidencia observada en el grupo control ( $ARB = AAB/I_0 = (I_e - I_0)/I_0$ ).
<b>Aumento relativo de riesgo (ARR)</b>	<i>Relative Risk Increase (RRI)</i>	<i>Medida del efecto</i> que se utiliza en los <i>ensayos clínicos aleatorios</i> cuando la intervención experimental aumenta el riesgo de una respuesta negativa o un acontecimiento adverso o indeseable. Expresa el aumento del riesgo obtenido en relación con el grupo control. Corresponde al cociente entre el aumento absoluto del riesgo y la incidencia o riesgo observada en el grupo control ( $ARR = AAR/I_0 = (I_e - I_0)/I_0$ ).

<b>Autonomía</b>	<i>Autonomy</i>	Capacidad de una persona para decidir por ella misma; está disminuida en los casos de ignorancia, inmadurez o incapacidad psíquica, cualquiera que sea su causa. El principio <i>bioético</i> de autonomía se considera sinónimo del de <i>respeto a las personas</i> . Ver también <i>consentimiento informado</i> .
<b>Azar</b>	<i>Random, chance</i>	Determinado por la suerte.
<b>Belmont, informe</b>	<i>Belmont report</i>	Informe elaborado por una Comisión del Congreso de Estados Unidos sobre la protección de los seres humanos objeto de experimentación biomédica y conductual, publicado en 1978. Este informe enumera los principios bioéticos básicos (ver <i>bioética</i> ).
<b>Beneficencia</b>	<i>Beneficence</i>	Principio <i>bioético</i> según el cual se debe tratar a las personas protegiéndolas de daños y asegurando su bienestar, lo que implica no hacer daño, maximizar el beneficio y minimizar los posibles riesgos e incomodidades, de forma que la relación beneficio/riesgo derivada de su participación en el estudio sea favorable.
<b>Biblioteca Cochrane</b>	<i>Cochrane Library</i>	Fuente de información electrónica creada por la <i>Colaboración Cochrane</i> , cuyo objetivo es proporcionar de forma rápida y simple las mejores evidencias necesarias para la toma de decisiones en la práctica clínica y sobre los efectos beneficiosos o perjudiciales de la atención sanitaria. Se trata de un conjunto de bases de datos publicado en CD-ROM actualizado trimestralmente. Incluye la Cochrane Database of Systematic Reviews, la Database of Abstracts of Reviews of Effectiveness, el Cochrane Controlled Trials Register y la Cochrane Review Methodology Database.
<b>Bioética</b>	<i>Bioethics, biomedical ethics</i>	Parte de la ética que trata de la moral y de las obligaciones de los seres humanos en el campo de la medicina y la investigación clínica y médico-biológica. Los principios bioéticos básicos destinados a proteger a los sujetos participantes en una investigación son los de <i>respeto por las personas</i> o <i>autonomía</i> , de <i>beneficencia</i> y de <i>justicia</i> . Algunos autores añaden el de <i>no maleficencia</i> , que distinguen del de <i>beneficencia</i> .
<b>Bland y Altman, método de</b>	<i>Bland-Altman's method</i>	Método gráfico para evaluar la <i>concordancia</i> y las diferencias entre dos métodos de medida de una <i>variable cuantitativa</i> . Se representan gráficamente las diferencias entre ambos métodos (eje de ordenadas) en relación con su media (eje de abscisas). Pueden representarse en relación con el verdadero valor de la variable cuando uno de los métodos puede considerarse el <i>estándar de referencia</i> . Si la concordancia fuera perfecta, los puntos se situarían exactamente sobre la línea correspondiente a la diferencia igual a 0. Este gráfico también permite observar si la magnitud de las diferencias entre los métodos es similar para todos los valores de la variable o si son mayores a medida que aumentan éstos. Si no varían, pueden analizarse las diferencias individuales: su media aritmética estima el sesgo relativo (diferencia sistemática entre ambos métodos) y su desviación estándar, el error aleatorio. Los límites del intervalo de dos desviaciones estándar alrededor de la media de las diferencias pueden considerarse los límites de concordancia entre ambas mediciones.

## Glosario

<b>Bonferroni, corrección de</b>	<i>Bonferroni's correction</i>	Procedimiento estadístico para realizar <i>comparaciones múltiples</i> . Se basa en utilizar, como <i>nivel de significación estadística</i> , el valor $0,05/k$ , donde $k$ es el número de comparaciones que se deben realizar.
<b>Buena Práctica Clínica (BPC)</b>	<i>Good Clinical Practices (GCP)</i>	Conjunto de normas administrativas que deben seguir los promotores, investigadores y monitores de los <i>ensayos clínicos</i> referentes a la forma en que son diseñados, realizados y comunicados, de modo que se asegure que los datos son fiables y sean aceptados por las agencias reguladoras de medicamentos, y que se protegen los derechos y la integridad de los sujetos, manteniendo la confidencialidad de los datos.
<b>Calibración</b>	<i>Calibration</i>	Proceso para establecer, con la mayor exactitud posible, la correspondencia entre las indicaciones de un instrumento de medida y los valores de magnitud que se miden con él.
<b>Causalidad</b>	<i>Causality, causation</i>	Juicio sobre la existencia de una relación causa-efecto entre el <i>factor de estudio</i> y la respuesta o enfermedad. Los criterios de causalidad de Hill se refieren a la <i>magnitud de la asociación, consistencia, especificidad, secuencia temporal, gradiente biológico, plausibilidad biológica, coherencia, evidencia experimental y analogía</i> . Ninguno de estos criterios es necesario ni suficiente para establecer una relación causal.
<b>Cegamiento</b>	<i>Blindness</i>	Ver <i>enmascaramiento</i> .
<b>Censurado</b>	<i>Censored</i>	En un <i>análisis de supervivencia</i> , sujetos en los que se desconoce su tiempo de supervivencia, ya sea porque al acabar el estudio aún no ha presentado el suceso de interés (retirado vivo) o porque ha interrumpido el seguimiento por cualquier motivo o ha fallecido por una causa no relacionada con el estudio (perdido).
<b>Ciego</b>	<i>Blind</i>	Ver <i>enmascaramiento</i> .
<b>Cociente</b>	<i>Quotient</i>	Resultado de la división de un numerador por un denominador.
<b>Cociente de probabilidad</b>	<i>Likelihood ratio</i>	Cociente entre la probabilidad de que una prueba diagnóstica proporcione un resultado determinado en un sujeto con la enfermedad y la de que lo proporcione en un sujeto sin la enfermedad. El cociente de probabilidad de un resultado positivo (CPP) corresponde a: CPP = sensibilidad/(1-especificidad). El cociente de probabilidad de un resultado negativo (CPN) corresponde a: CPN = (1 - sensibilidad)/especificidad. La utilidad clínica del CPP radica en que es una medida de la ganancia que se obtiene con un resultado positivo de una prueba diagnóstica aplicada a un determinado sujeto, gracias a la relación: <i>odds posprueba = odds preprueba × CPP</i> .
<b>Coeficiente de contingencia</b>	<i>Contingency coefficient</i>	Ver <i>contingencia, coeficiente de</i> .

<b>Coeficiente de correlación</b>	<i>Correlation coefficient</i>	1. Lineal de Pearson: <i>prueba estadística</i> que mide el grado de <i>asociación</i> entre dos <i>variables cuantitativas</i> . Los valores del coeficiente r oscilan entre -1 y +1. Un coeficiente de valor cero indica que no hay asociación lineal entre las dos variables, mientras que si es de -1 o +1 significa que la correlación es perfecta. Una correlación negativa significa que cuando los valores de una variable aumentan, los de la otra disminuyen. Una correlación positiva significa que cuando los valores de una variable aumentan, los de la otra también aumentan. La fuerza de la asociación depende del valor del coeficiente (ver <i>coeficiente de determinación</i> ). 2. No paramétricos: ver <i>Spearman</i> y <i>Kendall</i> .
<b>Coeficiente de correlación intraclass</b>	<i>Intraclass correlation coefficient</i>	Índice que sintetiza el grado de <i>concordancia</i> entre dos variables cuantitativas, que deriva del modelo del análisis de la varianza, y refleja la proporción de la variabilidad total que es debida a diferencias entre los sujetos. Se denomina R y sus valores van de 0 a 1. La concordancia suele considerarse muy buena si R es mayor de 0,90, buena si está entre 0,71 y 0,90, moderada entre 0,51 y 0,70, mediocre entre 0,31 y 0,50, y mala o nula si R es inferior a 0,31.
<b>Coeficiente de correlación parcial</b>	<i>Partial correlation coefficient</i>	<i>Coeficiente de correlación</i> de Pearson que mide la fuerza de la relación entre dos variables aleatorias eliminando el efecto de su común relación con una tercera variable.
<b>Coeficiente de determinación</b>	<i>Coefficient of determination</i>	Cuadrado del <i>coeficiente de correlación</i> . En un análisis de <i>correlación</i> , indica la proporción de la variabilidad conjunta entre las variables estudiadas. En un <i>análisis de regresión</i> , indica la proporción de la variabilidad de la <i>variable dependiente</i> que es explicada por la ecuación de regresión.
<b>Coeficiente de regresión</b>	<i>Regression coefficient</i>	Estimación del efecto de una <i>variable independiente</i> sobre una <i>variable dependiente</i> en un <i>modelo de regresión</i> . En la <i>regresión lineal simple</i> , corresponde a la pendiente de la recta de regresión. El coeficiente de regresión lineal de la variable X indica la magnitud del incremento de los valores de Y por cada aumento de una unidad en los valores de X. Cuando se trata de un modelo de <i>regresión múltiple</i> , este coeficiente está <i>ajustado</i> por el resto de variables independientes incluidas en la ecuación. Cuando se trata de una <i>regresión logística</i> , la variable dependiente es el logaritmo de la <i>odds</i> , por lo que su antilogaritmo (el número e elevado al valor del coeficiente) es una estimación de la <i>odds ratio</i> asociada a la variable X. Cuando la variable independiente está codificada como 0 (ausencia) y 1 (presencia), el coeficiente mide el efecto de la presencia de dicha variable en relación con su ausencia.
<b>Coeficiente de regresión ajustado</b>	<i>Adjusted regression coefficient</i>	<i>Coeficiente de regresión</i> en una ecuación de <i>regresión múltiple</i> , en cuya estimación se ha eliminado la influencia de uno o varios <i>factores de confusión</i> (términos incluidos en la ecuación).

## Glosario

<b>Coeficiente de variación (CV)</b>	<i>Coefficient of variation (CV)</i>	<i>Medida de dispersión</i> que permite comparar la variabilidad relativa entre dos series de valores. Se calcula dividiendo la <i>desviación estándar</i> por la <i>media aritmética</i> de cada serie, y multiplicando por 100. Equivale, por tanto, a la desviación estándar expresada en unidades de media, y se trata de una medida sin dimensiones e independiente de las unidades de medida.
<b>Coeficiente phi</b>	<i>Phi coefficient</i>	<i>Medida de asociación</i> en <i>tablas de contingencia</i> . Su valor oscila entre 0 (asociación nula) y 1 (asociación máxima), pudiendo interpretarse como un <i>coeficiente de correlación</i> .
<b>Coherencia</b>	<i>Coherence</i>	Criterio de <i>causalidad</i> que se refiere a que la relación observada no es discordante con el conocimiento actual sobre la historia natural y los mecanismos biológicos de la enfermedad.
<b>Cohorte</b>	<i>Cohort</i>	Grupo de sujetos con una característica o conjunto de características en común (generalmente la exposición al <i>factor de estudio</i> ), que son seguidos en el transcurso del tiempo. Puede tratarse de una generación, un grupo profesional, personas que hayan padecido una exposición determinada, portadores de una característica genética o habitantes en una comunidad definida geográficamente.
<b>Cohorte dinámica</b>	<i>Dynamic cohort</i>	<i>Cohorte</i> en la que los sujetos se van incluyendo a medida que progresá el estudio y son identificados, por lo que la fecha de inclusión puede ser diferente para cada sujeto. Opuesto a <i>cohorte fija</i> .
<b>Cohorte externa de comparación</b>	<i>External comparison cohort</i>	En un <i>estudio de cohorte</i> , grupo de comparación seleccionado por no estar expuesto al <i>factor de estudio</i> , pero que no procede de una misma cohorte general de la que ha sido extraída la cohorte expuesta, aunque se asume que representa adecuadamente a los sujetos no expuestos de la población de estudio. Ver también <i>cohorte interna de comparación</i> .
<b>Cohorte fija</b>	<i>Fixed cohort</i>	<i>Cohorte</i> formada por individuos que cumplen los criterios de inclusión en un momento determinado (tiempo cero), por lo que la fecha de inclusión o de inicio del estudio es la misma para todos los sujetos. Opuesto a <i>cohorte dinámica</i> .
<b>Cohorte interna de comparación</b>	<i>Internal comparison cohort</i>	En un <i>estudio de cohorte</i> , grupo de comparación formado por los sujetos de una cohorte general no expuestos al <i>factor de estudio</i> , mientras que los sujetos expuestos forman la cohorte de estudio. Dado que ambos grupos proceden de una misma cohorte general, la posibilidad de introducir <i>sesgos de selección</i> es menor que en la <i>cohorte externa de comparación</i> .
<b>Cohorte prospectiva</b>	<i>Prospective cohort</i>	<i>Cohorte</i> que es seguida en el tiempo, recogiendo los datos a medida que ocurren los hechos.
<b>Cohorte retrospectiva</b>	<i>Retrospective cohort</i>	<i>Cohorte</i> sobre la cual se obtiene información a partir de registros o archivos, ya que los hechos han ocurrido antes del desarrollo del estudio.

<b>Cointervención</b>	<i>Cointervention</i>	En un <i>ensayo clínico aleatorio</i> , aplicación de intervenciones adicionales diferentes a la de estudio en uno de los grupos o en ambos.
<b>Colaboración Cochrane</b>	<i>Cochrane Collaboration</i>	Organización internacional cuyo objetivo es ayudar a las personas a tomar decisiones sanitarias bien informadas a través de la elaboración, la revisión, el mantenimiento y la divulgación de <i>revisiones sistemáticas</i> sobre los beneficios y riesgos de las intervenciones sanitarias.
<b>Colinealidad</b>	<i>Collinearity</i>	Elevado grado de <i>correlación</i> o de relación lineal entre algunas de las <i>variables independientes</i> incluidas en un <i>modelo de regresión</i> , con la consecuencia de que los <i>errores estándar</i> de los <i>coeficientes de regresión</i> son muy grandes, e incluso de que las estimaciones de los coeficientes no sean plausibles y carezcan de utilidad.
<b>Comité Ético de Investigación Clínica (CEIC)</b>	<i>Institutional Review Board (IRB), Ethical Review Board (ERB)</i>	Comité encargado de evaluar la idoneidad de los <i>protocolos</i> de los <i>ensayos clínicos</i> con medicamentos y del equipo investigador propuesto, así como de la información escrita que se dará a los posibles sujetos de la investigación, la previsión de compensación en caso de lesiones y el tipo de <i>consentimiento</i> que otorgan los participantes. Además, debe realizar el seguimiento del ensayo clínico desde su inicio hasta su finalización.
<b>Comorbilidad</b>	<i>Comorbidity</i>	Enfermedades y problemas de salud que coexisten en un sujeto que participa en un estudio, además del proceso que es objeto de estudio.
<b>Comparabilidad</b>	<i>Comparability</i>	Grado de similitud entre los grupos que se comparan en un <i>estudio analítico</i> . Los grupos de estudio y de control no deben diferir significativamente en la distribución de las variables pronósticas de la respuesta o resultado, excepto en el factor de estudio. En los <i>ensayos clínicos aleatorios</i> , la <i>asignación aleatoria</i> tiende a asegurar esta comparabilidad, especialmente si el tamaño de la muestra es elevado.
<b>Comparaciones múltiples</b>	<i>Multiple comparison, multiple significance testing</i>	La aplicación de múltiples <i>pruebas estadísticas</i> para analizar unos datos aumenta la probabilidad de encontrar un resultado estadísticamente significativo simplemente por azar. Por ello, es necesario realizar una corrección del <i>nivel de significación estadística</i> que se utiliza o aplicar pruebas estadísticas de comparaciones múltiples, especialmente diseñadas para estas situaciones.
<b>Concordancia</b>	<i>Concordance, agreement</i>	Grado de correspondencia entre los valores de una variable proporcionados por dos métodos de medida u observadores diferentes, o bien entre medidas repetidas de la misma variable, un mismo método o un mismo observador. Cuando la variable es cualitativa, se mide mediante el índice kappa. Cuando es cuantitativa, mediante el coeficiente de correlación intraclass o el análisis de las diferencias individuales por el método de Bland y Altman.
<b>Concordancia interobservador</b>	<i>Interobserver agreement</i>	Grado en que dos o más observadores proporcionan resultados idénticos o muy similares al aplicar un instrumento de medida a una misma magnitud. Es un componente de la <i>fiabilidad</i> . Ver también <i>concordancia</i> .

## Glosario

<b>Concordancia intraobservador</b>	<i>Intraobserver agreement</i>	Grado en que un mismo observador proporciona resultados idénticos o muy similares al aplicar un instrumento de medida en más de una ocasión a una misma magnitud. Es un componente de la <i>fiabilidad</i> . Ver también <i>cordancia</i> .
<b>Confianza</b>	<i>Confidence</i>	En estadística, referida a los <i>intervalos de confianza</i> , valor que, multiplicado por 100, indica el porcentaje de intervalos del mismo tipo que incluyen realmente al parámetro objeto de estudio (habitualmente se utiliza el 95%). Garantía que se tiene de que la afirmación del intervalo sea correcta.
<b>Conformidad</b>	<i>Conformity</i>	Grado en que los resultados de una <i>prueba diagnóstica</i> coinciden con otra medida que no puede considerarse un patrón de referencia ( <i>patrón oro</i> ).
<b>Confusión</b>	<i>Confounding</i>	Fenómeno que aparece cuando la relación entre el factor de estudio y la variable de respuesta aparece distorsionada, de forma que puede ser total o parcialmente explicada por una tercera variable ( <i>factor de confusión</i> ), o bien puede quedar enmascarada por este factor.
<b>Consentimiento informado</b>	<i>Informed Consent</i>	Procedimiento formal por el que se solicita a los sujetos candidatos a participar en un estudio que acepten ser incluidos en el mismo. Debe reunir cuatro características: ser informado, comprendido, legalmente competente y voluntario.
<b>Consistencia</b>	<i>Consistency</i>	1. Criterio de <i>causalidad</i> que se refiere a la conformidad de los resultados de diferentes estudios realizados en lugares, poblaciones y tiempos diferentes. 2. Conformidad de los resultados observados en diferentes estratos (ver <i>análisis estratificado</i> y <i>modificación de efecto</i> ).
<b>Consistencia interna</b>	<i>Internal consistency</i>	Propiedad de un cuestionario que valora el grado en que los diferentes ítems o secciones del mismo miden un mismo atributo. Se evalúa mediante el alfa de Cronbach. Ver también <i>fiabilidad</i> .
<b>Contaminación</b>	<i>Contamination</i>	En los ensayos clínicos aleatorios, aplicación inadvertida de la intervención que es objeto de evaluación a las personas del grupo control, o incapacidad inadvertida de aplicar la intervención a las personas asignadas al grupo de estudio.
<b>Contingencia, coeficiente de</b>	<i>Contingency coefficient</i>	Medida de asociación en <i>tablas de contingencia</i> . Su valor mínimo es 0 (asociación nula), y su valor máximo es la raíz cuadrada del cociente $(q - 1)/q$ , donde $q$ es el menor del número de filas o columnas de la tabla.
<b>Contraste de hipótesis</b>	<i>Hypothesis testing</i>	Prueba estadística realizada para evaluar la plausibilidad de una hipótesis dada. El resultado es la probabilidad (valor $p$ ) de obtener el resultado encontrado, u otro más alejado de la hipótesis, si ésta fuera cierta. Cuando esta probabilidad es igual o menor al <i>nivel de significación estadística</i> prefijado, se rechaza la hipótesis.

<b>Control</b>	<i>Control</i>	1. Sujeto que forma parte del grupo de comparación ( <i>grupo control</i> ). En un <i>ensayo clínico aleatorio</i> , sujeto que recibe la intervención de referencia o el placebo. En un <i>estudio de casos y controles</i> , sujeto que no presenta el problema de salud de interés. 2. En estadística, controlar significa ajustar o tener en cuenta las influencias u observaciones externas (ver <i>ajuste</i> ).
<b>Controles concurrentes</b>	<i>Concurrent controls</i>	Grupo control cuyos datos han sido registrados en el mismo período de tiempo que el grupo de estudio.
<b>Controles históricos</b>	<i>Historical controls</i>	Grupo control cuyos datos fueron registrados en un período de tiempo anterior al del grupo de estudio. Teniendo en cuenta los cambios que se pueden producir a lo largo del tiempo en cuanto a diagnóstico, riesgo, pronóstico, etc., puede presentar diferencias sistemáticas en relación con el grupo de estudio e introducir un <i>sesgo</i> en la comparación.
<b>Corrección por continuidad</b>	<i>Continuity correction</i>	1. Artificio matemático que se realiza para mejorar la aproximación de una <i>variable discreta</i> por una <i>variable continua</i> . 2. Ver <i>Yates, corrección de</i> .
<b>Correlación</b>	<i>Correlation</i>	Grado en que dos variables varían conjuntamente. El coeficiente de correlación <i>r</i> de Pearson cuantifica el grado de relación lineal entre dos variables cuantitativas. Sus valores varían entre -1 (relación lineal negativa perfecta) y +1 (relación lineal positiva perfecta). Un valor igual a 0 indica la ausencia de relación lineal (ausencia de asociación o existencia de una relación no lineal). La proporción de variabilidad conjunta entre ambas variables viene cuantificada por el cuadrado de <i>r</i> , conocido como coeficiente de determinación. Ver también <i>asociación</i> . Correlación no es sinónimo de <i>concordancia</i> .
<b>Covarianza</b>	<i>Covariance</i>	Medida de la variación conjunta de dos variables. Es sensible a las unidades de medida.
<b>Cramer, prueba de la V de</b>	<i>Cramer's V test</i>	Medida de asociación en tablas de contingencia, basada en la distribución de la <i>ji al cuadrado</i> . En tablas $2 \times 2$ coincide con el <i>coeficiente phi</i> .
<b>Cuartil</b>	<i>Quartile</i>	Medida de posición que corresponde a los <i>percentiles</i> 25, 50 y 75. Por ejemplo, el segundo cuartil corresponde al valor que deja por debajo el 50% de las observaciones, una vez ordenadas de menor a mayor, y equivale a la <i>mediana</i> . La diferencia entre el tercer y el primer cuartil corresponde al <i>recorrido intercuartílico</i> .
<b>Cumplimiento</b>	<i>Compliance, adherence</i>	Grado en que un paciente sigue el tratamiento prescrito, sea farmacológico o no, y las instrucciones que se le han dado en un estudio. Sinónimos: adherencia, observancia.
<b>Curtosis</b>	<i>Kurtosis</i>	Grado de apuntamiento de una distribución. Se utiliza para determinar la desviación de la normalidad de la misma.

## Glosario

<b>Curva ROC</b>	<i>Receiver Operating Characteristic (ROC) Curve</i>	Gráfico que se utiliza cuando los valores de una <i>prueba diagnóstica</i> se miden en una escala cuantitativa. En el eje de ordenadas se sitúa la <i>sensibilidad</i> (proporción de verdaderos positivos) y en el de abscisas, el complementario de la <i>especificidad</i> ( $1 - \text{especificidad} = \text{proporción de falsos positivos}$ ). En este gráfico se representan los puntos correspondientes a los diferentes <i>puntos de corte</i> y se unen para dibujar una curva. El área bajo la curva indica la probabilidad de clasificar correctamente un par de individuos sano y enfermo, seleccionados al azar, al aplicar la prueba diagnóstica.
<b>Datos primarios</b>	<i>Primary data</i>	Información referente a los sujetos incluidos en un estudio que se recoge específicamente para la realización de la investigación.
<b>Datos secundarios</b>	<i>Secondary data</i>	Información referente a los sujetos incluidos en un estudio que se obtiene de registros previos o fuentes documentales existentes, y que fueron recogidos con un propósito diferente al de la investigación (p. ej., estadísticas de mortalidad, censo poblacional, historias clínicas, etc.).
<b>Decil</b>	<i>Decile</i>	<i>Medida de posición</i> que corresponde a los <i>percentiles</i> 10, 20, 30, 40, ..., 90. Por ejemplo, el decil 4 corresponde al percentil 40, y corresponde al valor que deja por debajo el 40% de las observaciones, una vez ordenadas de menor a mayor.
<b>Densidad de incidencia</b>	<i>Incidence density</i>	Tasa que mide la frecuencia de aparición de un suceso o enfermedad. Se define como el número de nuevos casos que aparecen por unidad de tiempo-población (habitualmente años-persona) de observación.
<b>Descriptivo</b>	<i>Descriptive</i>	Ver <i>estudio descriptivo</i> .
<b>Desviación</b>	<i>Deviation</i>	En estadística, diferencia entre el valor de una de las observaciones de una serie y la <i>media aritmética</i> de la misma.
<b>Desviación estándar (DE)</b>	<i>Standard deviation (SD)</i>	<i>Medida de dispersión</i> que corresponde a la raíz cuadrada de la <i>varianza</i> . Si la variable sigue una <i>distribución normal</i> , el intervalo de una desviación estándar alrededor de la media incluye aproximadamente el 68% de las observaciones, y el intervalo de dos desviaciones estándar alrededor de la media incluye aproximadamente el 95% de las observaciones.
<b>Desviación media</b>	<i>Mean deviation</i>	En estadística, media aritmética de las <i>desviaciones</i> de los valores de la serie respecto a su <i>media</i> .
<b>Desviación típica</b>	<i>Standard deviation (SD)</i>	Ver <i>desviación estándar</i> .
<b>Diagrama de barras</b>	<i>Bar chart</i>	Representación gráfica de la <i>distribución de frecuencias</i> de una <i>variable cualitativa</i> , en la que la <i>frecuencia</i> de cada categoría se representa mediante una barra, cuya longitud es proporcional al número de casos. Las barras pueden ser tanto horizontales como verticales. Cuando se trata de <i>variables nominales</i> , dado que el orden de las categorías es arbitrario, suelen ordenarse de mayor a menor frecuencia de observación, lo que facilita su comprensión.

<b>Diagrama de sectores</b>	<i>Pie chart</i>	Representación gráfica de la frecuencia de las categorías de una <i>variable cualitativa</i> , que consiste en un círculo, dentro del cual se asigna a cada categoría un sector de área proporcional a su frecuencia.
<b>Diferencia de incidencias (DI)</b>	<i>Incidence difference</i>	Medida del efecto del factor de estudio que corresponde a la diferencia de las <i>incidencias</i> observadas en el grupo de estudio y el grupo control ( $DI = I_c - I_0$ ). Se considera una medida de impacto, que representa la cantidad de la incidencia que puede ser atribuida al factor de estudio (exceso de riesgo).
<b>Diferencia de riesgos</b>	<i>Risk difference</i>	Ver <i>diferencia de incidencias</i> .
<b>Diferencia estandarizada de medias</b>	<i>Standardised mean difference</i>	Medida del efecto que corresponde a la diferencia entre dos medias dividida por una estimación de la desviación estándar interna del grupo. Esta medida estandarizada, al carecer de unidades, permite comparar resultados medidos de manera diferente en los diferentes estudios incluidos en una <i>revisión sistemática</i> .
<b>Diferencia ponderada de medias</b>	<i>Weighted mean difference</i>	Método utilizado en un <i>metaanálisis</i> para combinar medidas de escalas continuas cuando la media, la desviación estándar y el tamaño de la muestra de cada grupo son conocidos. Asume que todos los estudios han medido el resultado en la misma escala. El peso de cada estudio viene dado por la <i>precisión</i> del estimador del efecto.
<b>Diseño cruzado</b>	<i>Cross-over design</i>	Ver <i>ensayo clínico cruzado</i> .
<b>Diseño factorial</b>	<i>Factorial design</i>	Ensayo clínico aleatorio en que los sujetos son asignados a una combinación de dos o más intervenciones o a una combinación de diferentes dosis de un mismo tratamiento. Se utiliza para realizar diferentes <i>contrastos de hipótesis</i> en un mismo ensayo o para evaluar la existencia de interacción entre tratamientos.
<b>Diseño híbrido</b>	<i>Hybrid design</i>	Diseño que comparte características tanto de los <i>estudios de cohortes</i> como de los <i>estudios de casos y controles</i> , obviando algunas de las desventajas de estos últimos. Se caracteriza por analizar todos los casos aparecidos en una cohorte seguida en el tiempo, y utilizar como <i>controles</i> sólo una muestra de los sujetos de esa misma cohorte. Los diseños híbridos más importantes son los <i>estudios de casos y controles anidados en una cohorte</i> y los <i>estudios de cohorte y casos</i> .
<b>Diseño paralelo</b>	<i>Parallel design</i>	Estudio en que los grupos son seguidos de forma simultánea (ver <i>controles concurrentes</i> ).
<b>Diseño secuencial</b>	<i>Sequential design</i>	Ensayo clínico en que el <i>tamaño de la muestra</i> no está pre-determinado y el análisis de las observaciones se realiza a medida que se dispone de ellas, de forma que se van incluyendo sujetos hasta que la diferencia acumulada entre los tratamientos alcanza unos límites definidos clínica y estadísticamente, o hasta que se comprueba que es improbable que se observe una diferencia. Se realiza cuando el tiempo de seguimiento no es largo y la evaluación de la respuesta puede hacerse con rapidez.

## Glosario

<b>Dispersión, medidas de</b>	<i>Dispersion measures</i>	Medidas que expresan las diferencias o la variabilidad entre los valores de una <i>variable cuantitativa</i> medidos en una serie de observaciones. Ver <i>recorrido, recorrido intercuartílico, desviación estándar y varianza</i> .
<b>Distribución</b>	<i>Distribution</i>	Ver <i>distribución de probabilidad</i> .
<b>Distribución binomial</b>	<i>Binomial distribution</i>	<i>Distribución de probabilidad</i> de una <i>variable discreta</i> , que es el número de individuos de una muestra que tienen un determinado valor de una <i>variable dicotómica</i> (p. ej., número de pacientes que mejoran de entre los 100 que reciben un tratamiento).
<b>Distribución de frecuencias</b>	<i>Frequency distribution</i>	Conjunto de valores o intervalos de valores de una variable, junto con las <i>frecuencias</i> correspondientes de cada uno de ellos.
<b>Distribución de Poisson</b>	<i>Poisson distribution</i>	<i>Distribución de probabilidad</i> de una <i>variable discreta</i> aplicable al número de veces que un suceso ocurre. Se utiliza para determinar la probabilidad de que un suceso raro ocurra un determinado número de veces, cuando el número de ensayos es muy grande y la probabilidad muy pequeña; por ejemplo, el número de hospitalizaciones en un período de tiempo determinado.
<b>Distribución de probabilidades</b>	<i>Probability distribution</i>	Función que proporciona las probabilidades de que una variable sea igual a cada uno de los posibles valores de una secuencia. Son ejemplos la <i>distribución normal</i> , la <i>distribución binomial</i> , la <i>distribución de Poisson</i> , la distribución de la <i>t al cuadrado</i> o la distribución de la <i>t de Student</i> .
<b>Distribución normal</b>	<i>Normal distribution</i>	<i>Distribución de probabilidad</i> de una <i>variable continua</i> que tiene forma acampanada (campana de Gauss), es simétrica, es asintótica al eje de abscisas (cualquier valor entre -infinito y +infinito es teóricamente posible), la <i>media</i> , <i>mediana</i> y <i>moda</i> coinciden, y la distancia entre la media y el punto de inflexión corresponde a la <i>desviación estándar</i> (cuanto menor es ésta, más apuntada es la curva). Por tanto, la distribución normal queda perfectamente definida por sus parámetros media y desviación estándar.
<b>Doble ciego</b>	<i>Double blind</i>	Procedimiento de <i>enmascaramiento</i> que suele utilizarse en los <i>ensayos clínicos aleatorios</i> , de manera que ni los pacientes ni el investigador o el personal clínico en contacto con el paciente conocen el grupo al que han sido asignados. Cuando se comparan fármacos, requiere que su forma de presentación sea idéntica.
<b>Doble simulación</b>	<i>Double dummy</i>	Procedimiento de <i>enmascaramiento</i> que se utiliza en los <i>ensayos clínicos aleatorios</i> , cuando se comparan fármacos con formas de presentación o vías de administración diferentes. Requiere disponer de <i>placebos</i> adecuados de ambos tratamientos, de forma que cada sujeto recibe la forma activa de uno de ellos y el placebo del otro, en función del grupo al que ha sido asignado.
<b>Dosis-respuesta</b>	<i>Dose response</i>	Ver <i>gradiente dosis-respuesta</i> .
<b>EEM</b>	<i>SEM</i>	Acrónimo de <i>error estándar de la media</i> .

<b>Efectividad</b>	<i>Effectiveness</i>	Grado en que una determinada intervención produce un resultado beneficioso en condiciones reales de aplicación.
<b>Efecto</b>	<i>Effect</i>	Resultado de una causa, acción, intervención o factor de estudio.
<b>Efecto de arrastre</b>	<i>Carry-over effect</i>	Ver <i>efecto residual</i> .
<b>Efecto Hawthorne</b>	<i>Hawthorne effect</i>	Respuesta inducida por el conocimiento de los participantes de que están siendo estudiados. En los estudios sin <i>grupo control</i> no se puede separar este efecto del debido al factor de estudio, por lo que éste no puede estimarse de una forma válida.
<b>Efecto período</b>	<i>Period effect</i>	En un <i>ensayo clínico cruzado</i> , existencia de diferencias en los efectos de los tratamientos debidas al diferente período de tiempo en que son administrados.
<b>Efecto placebo</b>	<i>Placebo effect</i>	Respuesta que se produce en un paciente como consecuencia de la administración de un tratamiento, pero que no puede considerarse un efecto específico del mismo. En los estudios que evalúan la eficacia de un tratamiento, si no existe un <i>grupo control</i> que reciba una intervención similar a la de estudio en todas las características excepto en la actividad farmacológica, no se puede separar este efecto del debido al tratamiento.
<b>Efecto residual</b>	<i>Carry-over effect</i>	Persistencia del efecto después de interrumpir el tratamiento. Cuando en un estudio un sujeto recibe más de un tratamiento, debe incluirse un <i>período de lavado</i> que elimine este efecto. Ver también <i>ensayo clínico cruzado</i> .
<b>Efecto secuencia</b>	<i>Sequence effect</i>	En un <i>ensayo clínico cruzado</i> , existencia de diferencias en los efectos de los tratamientos debidas al orden o la secuencia en que son administrados.
<b>Eficacia</b>	<i>Efficacy</i>	Grado en que una determinada intervención produce un resultado beneficioso en condiciones ideales de aplicación. Idealmente se determina según los resultados de un <i>ensayo clínico aleatorio</i> .
<b>Eficiencia</b>	<i>Efficiency</i>	Grado en que una determinada intervención produce un resultado beneficioso, en relación con el esfuerzo empleado (recursos humanos y materiales, costes económicos, tiempo, etc.).
<b>Emparejamiento</b>	<i>Matching</i>	Proceso por el que se selecciona, para cada uno de los casos, uno o más controles que se asemejan a él en determinadas variables (p. ej., grupo de edad y sexo), con la finalidad de controlar potenciales <i>factores de confusión</i> , haciendo que los grupos de estudio y de control sean comparables por dichas variables.
<b>Enmascaramiento</b>	<i>Blindness</i>	Procedimiento utilizado con el fin de que los miembros del equipo investigador, los participantes en el estudio o los profesionales que realizan las mediciones no conozcan algunos hechos u observaciones que pudieran ejercer un cambio en sus acciones o decisiones y sesgar los resultados. Ver <i>simple ciego</i> , <i>doble ciego</i> , <i>triple ciego</i> y <i>evaluación ciega</i> .

## Glosario

<b>Ensayo clínico</b>	<i>Clinical trial</i>	Estudio que implica la administración de una intervención (habitualmente farmacológica) en seres humanos, con la finalidad de evaluar su eficacia y seguridad. El uso de este término es variable, desde los estudios de la primera aplicación de un medicamento en seres humanos sin un grupo control para determinar sus efectos farmacodinámicos y su perfil farmacocinético, hasta los ensayos clínicos aleatorios. Habitualmente se refiere a los ensayos clínicos controlados, sean aleatorios o no.
<b>Ensayo clínico abierto</b>	<i>Open clinical trial, non blinded clinical trial</i>	Habitualmente se refiere a un <i>ensayo clínico</i> que no utiliza <i>enmascaramiento</i> .
<b>Ensayo clínico aleatorio</b>	<i>Randomized Clinical Trial</i>	<i>Ensayo clínico</i> en que se evalúa la eficacia de una intervención, comparando habitualmente dos grupos de sujetos de forma concurrente, uno de los cuales recibe la intervención de estudio y el otro la de referencia, un placebo o ninguna intervención, en el que la asignación de los sujetos a los grupos se realiza de forma aleatoria ( <i>asignación aleatoria</i> ).
<b>Ensayo clínico controlado</b>	<i>Controlled clinical trial</i>	<i>Ensayo clínico</i> que compara uno o más grupos de estudio con uno o más grupos control o de comparación. La <i>asignación</i> a los distintos grupos puede ser <i>aleatoria</i> o no.
<b>Ensayo clínico cruzado</b>	<i>Cross-over clinical trial</i>	<i>Ensayo clínico aleatorio</i> en que los sujetos reciben dos o más intervenciones en períodos sucesivos y en una secuencia que ha sido determinada al azar, de modo que cada sujeto actúa como su propio control. Para evitar el <i>efecto residual</i> de la intervención administrada en el primer período, se incluye un <i>período de lavado</i> entre los períodos lo suficientemente largo para que dicho efecto desaparezca.
<b>Ensayo clínico explicativo</b>	<i>Explicative clinical trial</i>	<i>Ensayo clínico</i> diseñado con la finalidad de adquirir conocimientos científicos y explicaciones biológicas sobre <i>eficacia</i> , para lo que utiliza criterios de selección restrictivos y se realiza en condiciones también restrictivas, ideales para poner de manifiesto el efecto que se estudia, pero muy diferentes de las de la práctica habitual. Opuesto a <i>ensayo clínico pragmático</i> .
<b>Ensayo clínico no controlado</b>	<i>Non-controlled clinical trial</i>	<i>Ensayo clínico</i> en el que no existe un grupo de <i>control concurrente</i> . Ver también <i>estudios antes-después</i> .
<b>Ensayo clínico pragmático</b>	<i>Pragmatic clinical trial</i>	<i>Ensayo clínico</i> diseñado con la finalidad de medir la <i>efectividad</i> de una intervención en su aplicación clínica habitual, para lo que utiliza criterios de selección laxos y se realiza en condiciones lo más parecidas posibles a las de la práctica habitual. Opuesto a <i>ensayo clínico explicativo</i> .
<b>Ensayo comunitario</b>	<i>Community trial</i>	Ensayo realizado en comunidades, evaluando los resultados observados en global en cada una de ellas. Es el diseño apropiado para la evaluación de intervenciones que no pueden ser asignadas de forma individual.
<b>Ensayo factorial</b>	<i>Factorial design</i>	Ver <i>diseño factorial</i> .
<b>Ensayo secuencial</b>	<i>Sequential trial</i>	Ver <i>diseño secuencial</i> .

<b>Equivalencia</b>	<i>Equivalence</i>	Ver <i>estudio de equivalencia</i> .
<b>Error aleatorio</b>	<i>Random error</i>	Error debido al azar. Se debe tanto al hecho de que se trabaja con <i>muestras</i> de individuos como a la variabilidad inherente al proceso de medición de variables.
<b>Error alfa</b>	<i>Alpha error</i>	En una prueba de <i>contraste de hipótesis</i> , conclusión de que existe una diferencia o asociación cuando de hecho no la hay (rechazo de la <i>hipótesis nula</i> cuando es cierta). La probabilidad de cometer este tipo de error viene cuantificada por el valor <i>p</i> . Es un concepto similar al de probabilidad de un resultado falso positivo en una prueba diagnóstica. También se denomina error tipo I.
<b>Error beta</b>	<i>Beta error</i>	En una prueba de <i>contraste de hipótesis</i> , conclusión de que no existe una diferencia o asociación cuando en realidad existe (no rechazo de la <i>hipótesis nula</i> cuando es falsa). Es un concepto similar al de probabilidad de un resultado falso negativo en una prueba diagnóstica. También se denomina error tipo II. Su complementario ( $1 - \beta$ ) se conoce como <i>potencia estadística</i> .
<b>Error estándar (EE)</b>	<i>Standard error (SE)</i>	Desviación estándar de una <i>estimación</i> . Es una medida del grado de incertidumbre con que una estimación obtenida en una muestra se acerca al verdadero valor en la población ( <i>precisión</i> ). Ver también <i>intervalo de confianza</i> .
<b>Error sistemático</b>	<i>Systematic error, bias</i>	Ver <i>sesgo</i> .
<b>Error tipo I</b>	<i>Type I error</i>	Ver <i>error alfa</i> .
<b>Error tipo II</b>	<i>Type II error</i>	Ver <i>error beta</i> .
<b>Especificidad</b>	<i>Specificity</i>	1. Cuando se refiere a pruebas diagnósticas, <i>proporción</i> de individuos no enfermos que son correctamente clasificados como tales por una <i>prueba diagnóstica</i> (resultado de la prueba negativo). Una prueba muy específica da pocos resultados falsos positivos. 2. Criterio de <i>causalidad</i> que se refiere al grado en que un solo factor se asocia con un único efecto.
<b>Estadística</b>	<i>Statistics</i>	Conjunto de métodos para recoger, clasificar, representar y resumir datos, así como para extraer conclusiones científicas a partir de ellos. Ver <i>estadística descriptiva</i> y <i>estadística inferencial</i> .
<b>Estadística descriptiva</b>	<i>Descriptive statistics</i>	Procedimientos para la recogida, organización, presentación y síntesis de datos de una manera científica.
<b>Estadística inferencial</b>	<i>Inferential statistics</i>	Procedimientos mediante los cuales se establecen conclusiones relacionadas con <i>poblaciones</i> (inferencias) a partir de los resultados obtenidos en <i>muestras</i> .
<b>Estadísticamente no significativo</b>	<i>Statistically non-significant</i>	Un resultado o prueba lo es cuando no se rechaza la <i>hipótesis nula</i> . Significa que no se dispone de la suficiente evidencia para descartar que la diferencia o asociación observada pueda ser debida al azar (en el supuesto de que no existiera en la realidad), pero no es sinónimo de ausencia de diferencia o asociación. Habitualmente es sinónimo de $p > 0,05$ (si el <i>nivel de significación estadística</i> se ha establecido en el 5%).

## Glosario

<b>Estadísticamente significativo</b>	<i>Statistically significant</i>	Un resultado o prueba lo es cuando se rechaza la <i>hipótesis nula</i> y se acepta la <i>hipótesis alternativa</i> . Significa que se dispone de la suficiente evidencia para concluir que la diferencia o asociación observada no se debe al azar (es improbable que se hubiera observado si no existiera en la realidad). Habitualmente, sinónimo de $p < 0,05$ (si el <i>nivel de significación estadística</i> se ha establecido en el 5%).
<b>Estadístico</b>	<i>Statistic</i>	Función de los valores de una muestra.
<b>Estandarización</b>	<i>Standardization</i>	1. Procedimientos utilizados para eliminar en lo posible los efectos de las diferencias en la estructura de edad o la distribución de otros <i>factores de confusión</i> cuando se comparan dos o más poblaciones. Se basan en una ponderación de las tasas específicas por sexo, grupos de edad u otras variables de acuerdo con una distribución determinada de estas variables. Ver también <i>ajuste</i> . 2. De una variable: ver <i>tipificación</i> .
<b>Estimación</b>	<i>Estimate</i>	Medida de la magnitud de una cantidad, de la que se sabe, se cree o se sospecha que incorpora cierto grado de error. Ver <i>estimador</i> .
<b>Estimación cruda</b>	<i>Raw estimate</i>	Estimación de un efecto o asociación sin tener en cuenta los potenciales <i>factores de confusión</i> . Opuesta a <i>estimación ajustada</i> .
<b>Estimación ajustada</b>	<i>Adjusted estimate</i>	Estimación de un efecto o asociación teniendo en cuenta los potenciales <i>factores de confusión</i> (ver <i>ajuste</i> ). Opuesta a <i>estimación cruda</i> .
<b>Estimación puntual</b>	<i>Point estimate</i>	Valor observado en la muestra estudiada, utilizado como la mejor estimación disponible del verdadero valor del parámetro de interés en la población de referencia de la que procede la muestra. Esta estimación no tiene en cuenta la variabilidad inherente al <i>muestreo</i> , por lo que es necesario acompañarla de una medida de la <i>precisión</i> con que se ha obtenido (ver <i>error estándar</i> e <i>intervalo de confianza</i> ).
<b>Estimador</b>	<i>Estimator</i>	Medida resumen calculada en una <i>muestra</i> que se utiliza para realizar <i>inferencias</i> sobre la población. Ver <i>estimación puntual</i> e <i>intervalo de confianza</i> .
<b>Estocástico</b>	<i>Stochastic</i>	Ver <i>aleatorio</i> .
<b>Estratificación</b>	<i>Stratification</i>	Procedimiento por el que se distribuyen los sujetos de una muestra en submuestras o estratos, definidas por las diferentes categorías de una o varias variables. La utilización de un <i>análisis estratificado</i> permite controlar el efecto de factores de <i>confusión</i> y evaluar la existencia de <i>modificación de efecto</i> .
<b>Estudio analítico</b>	<i>Analytic study</i>	Estudio diseñado para evaluar una presunta <i>relación causal</i> entre un <i>factor de estudio</i> (una exposición, una intervención, etc.) y una <i>variable de respuesta</i> . Los diseños analíticos más frecuentes son los <i>estudios de casos y controles</i> , los <i>estudios de cohortes</i> y los <i>ensayos clínicos aleatorios</i> .
<b>Estudio antes-después</b>	<i>Before-after study</i>	Estudio de intervención sin grupo control en el que se evalúa la respuesta de un grupo de sujetos antes y después de una determinada intervención.

<b>Estudio de asociación cruzada</b>	<i>Cross-association study</i>	Diseño <i>transversal</i> que se utiliza para estudiar la relación entre dos o más variables en una población definida en un momento determinado del tiempo. Dado que las variables se miden simultáneamente, existe una ambigüedad temporal que dificulta que pueda establecerse una <i>secuencia temporal</i> adecuada, por lo que su utilidad para evaluar una <i>relación causal</i> es limitada.
<b>Estudio de casos y controles</b>	<i>Case-control study</i>	Diseño en que se seleccionan dos grupos de individuos, uno con el resultado de interés (casos) y otro sin él (controles), midiendo «hacia atrás» su exposición previa y observando si hay diferencias entre ellos.
<b>Estudio de casos y controles anidados en una cohorte</b>	<i>Nested case-control study</i>	Diseño <i>híbrido</i> en que, a partir de un <i>estudio de cohortes</i> ya realizado o que se está llevando a cabo, se identifican todos los sujetos que han desarrollado la enfermedad, que constituyen el grupo de casos. En el momento en que aparece un caso, se selecciona aleatoriamente uno o varios controles entre los sujetos en situación de riesgo en ese momento.
<b>Estudio de cohorte y casos</b>	<i>Case-cohort study</i>	Diseño <i>híbrido</i> en que, a partir de un <i>estudio de cohortes</i> que se está llevando a cabo, se selecciona una muestra o subcohorte que servirá de grupo de comparación para todos los casos que aparezcan durante el seguimiento, con independencia de si ya pertenecían o no a la subcohorte.
<b>Estudio de cohortes</b>	<i>Cohort study</i>	Diseño en que se seleccionan dos grupos o cohortes de individuos, uno con la exposición de interés y otro sin ella, que son seguidos en el tiempo para determinar si hay diferencias en la aparición del resultado.
<b>Estudio de equivalencia</b>	<i>Equivalence study, equivalence trial</i>	Ensayo clínico cuyo objetivo es evaluar si una nueva intervención tiene la misma eficacia que otra ya conocida. Obliga a definir, antes de su realización, la mínima diferencia entre las intervenciones a partir de la cual se considerará que existen diferencias.
<b>Estudio de prevalencia</b>	<i>Prevalence study</i>	Diseño <i>transversal</i> cuya finalidad es la estimación de la frecuencia de una enfermedad o característica en una población definida en un momento determinado del tiempo ( <i>prevalencia</i> ).
<b>Estudio descriptivo</b>	<i>Descriptive study</i>	Estudios diseñados solamente con la finalidad de describir la distribución existente de una o más variables, sin realizar <i>contraste de hipótesis</i> ni evaluar presuntas <i>relaciones causales</i> . Los diseños descriptivos más frecuentes son los <i>estudios transversales</i> , el seguimiento de una <i>cohorte</i> de sujetos y las <i>series de casos</i> .
<b>Estudio ecológico</b>	<i>Ecological study</i>	Diseño en que las unidades de análisis son agregaciones de individuos, a menudo basadas en criterios geográficos o temporales. Suelen utilizarse para comparar la frecuencia de un problema de salud en diferentes áreas geográficas o su variación a lo largo del tiempo en una comunidad determinada.

## Glosario

<b>Estudio negativo</b>	<i>Negative study</i>	Término utilizado para referirse a los estudios que no alcanzan resultados estadísticamente significativos. El término puede prestarse a confusión, ya que se refiere simultáneamente a la significación estadística y a la dirección del efecto.
<b>Estudio positivo</b>	<i>Positive study</i>	Término utilizado para referirse a los estudios que alcanzan resultados estadísticamente significativos, mostrando un efecto beneficioso de la intervención estudiada o un efecto pernicioso cuando se evalúa riesgo o efectos secundarios. El término puede prestarse a confusión, ya que se refiere simultáneamente a la significación estadística y a la dirección del efecto.
<b>Estudio transversal</b>	<i>Cross-sectional study</i>	Diseño en que se estudia una muestra de sujetos en un punto del tiempo o un intervalo, determinando simultáneamente la presencia de la exposición y del resultado.
<b>Evaluación ciega</b>	<i>Blind evaluation</i>	Procedimiento de <i>enmascaramiento</i> por el que el observador que mide la variable de respuesta desconoce el grupo al que pertenece el sujeto de estudio.
<b>Evidencia experimental</b>	<i>Experimental evidence</i>	Criterio de <i>causalidad</i> que se refiere a la demostración mediante métodos experimentales de un mecanismo de acción que explique los hechos observados.
<b>Exactitud</b>	<i>Validity</i>	Ver <i>validez</i> .
<b>Exceso de riesgo</b>	<i>Risk excess</i>	Ver <i>diferencia de incidencias</i> .
<b>Exclusiones</b>	<i>Exclusions, pre-assignment exclusions</i>	En un <i>ensayo clínico aleatorio</i> , sujetos candidatos que no son finalmente incluidos en el estudio porque no cumplen todos los criterios de selección o no dan su <i>consentimiento informado</i> para participar en el estudio. Dado que se producen antes de la <i>asignación aleatoria</i> , no afectan a la <i>validez interna</i> de los resultados, pero pueden limitar su <i>validez externa</i> .
<b>Experimental</b>	<i>Experimental</i>	Diseño en que el equipo investigador controla el <i>factor de estudio</i> , en el sentido de que decide quién lo recibe y lo administra de forma deliberada para la realización de una investigación, según un plan preestablecido. El diseño experimental más importante es el <i>ensayo clínico aleatorio</i> . Opuesto a <i>observacional</i> .
<b>Extrapolación</b>	<i>Extrapolation</i>	Aplicación de los resultados de un estudio a una población diferente de la estudiada. Muy relacionada con la <i>validez externa</i> .
<b>F de Snedecor, prueba de la</b>	<i>F test</i>	Prueba estadística para la comparación de dos <i>varianzas</i> en el caso de muestras independientes.
<b>Factor de confusión</b>	<i>Confounding factor</i>	Factor que distorsiona la verdadera relación entre el factor de estudio y la variable de respuesta (ver <i>confusión</i> ). Debe ser un factor pronóstico de la respuesta, estar asociado al factor de estudio y no ser un paso intermedio en la cadena causal entre el factor de estudio y la variable de respuesta.

<b>Factor de estudio</b>	<i>Study factor</i>	Variable (exposición, intervención terapéutica o preventiva, etc.) cuyo efecto o asociación con una <i>variable de respuesta</i> se desea estimar en un estudio.
<b>Factor de riesgo</b>	<i>Risk factor</i>	Conducta o hábito de vida, exposición o característica congénita o heredada que, según la evidencia epidemiológica disponible, se conoce que está asociada con el <i>riesgo</i> de desarrollar una enfermedad o problema de salud.
<b>Factor pronóstico</b>	<i>Prognostic factor</i>	Característica de un sujeto (sociodemográfica, de la enfermedad, comorbilidad, etc.) que puede predecir los eventuales resultados, tanto positivos como negativos, en el curso de una enfermedad, sin que necesariamente deba existir una <i>relación causal</i> entre ellos.
<b>Falacia ecológica</b>	<i>Ecological fallacy</i>	Elaboración inadecuada de conclusiones a nivel individual a partir de datos procedentes de <i>estudios ecológicos</i> .
<b>Fiabilidad</b>	<i>Reliability</i>	Grado en que un instrumento de medida proporciona valores idénticos o muy similares cuando se aplica en más de una ocasión. Suele considerarse sinónimo de <i>precisión</i> . Se evalúa analizando la <i>concordancia</i> entre las diferentes mediciones obtenidas. Tiene diferentes componentes: <i>repetibilidad</i> (fiabilidad test-retest), <i>concordancia intraobservador</i> y <i>concordancia interobservador</i> . Cuando se refiere a cuestionarios, se incluye un componente adicional: la <i>consistencia interna</i> .
<b>Fiabilidad interobservador</b>	<i>Interobserver agreement</i>	Ver <i>concordancia interobservador</i> .
<b>Fiabilidad intraobservador</b>	<i>Intraobserver agreement</i>	Ver <i>concordancia intraobservador</i> .
<b>Fiabilidad test-retest</b>	<i>Test-retest reliability</i>	Ver <i>repetibilidad</i> .
<b>Fisher, prueba exacta de</b>	<i>Fisher's exact test</i>	Prueba estadística para la comparación de dos proporciones con datos <i>independientes</i> , en cualquier circunstancia. Habitualmente se utiliza cuando no se cumplen las condiciones de aplicación de la prueba de la <i>ji al cuadrado</i> ni de la <i>corrección de Yates</i> .
<b>Fracción atribuible</b>	<i>Attributable fraction</i>	Medida de impacto que corresponde a una <i>diferencia de incidencias</i> relativa. La fracción atribuible en los expuestos corresponde al cociente entre la diferencia de incidencias y la incidencia del grupo expuesto ( $FAE = [I_e - I_o]/I_e$ ; $FAE = [RR - 1]/RR$ ). Para calcular la fracción atribuible poblacional, la incidencia del grupo expuesto se sustituye por la poblacional ( $FAP = [I_p - I_o]/I_p$ ; $FAP = FAE \times \text{proporción de exposición entre los enfermos}$ ).
<b>Fracción etiológica</b>	<i>Etiologic fraction</i>	Ver <i>fracción atribuible</i> .
<b>Frecuencia</b>	<i>Frequency</i>	En estadística, número de veces en que se ha presentado un determinado valor en una serie de observaciones (frecuencia absoluta). Cuando se divide por el número de total de observaciones, se obtiene la frecuencia relativa de presentación de dicho valor en la serie.
<b>Fuerza de la asociación</b>	<i>Strength of the association</i>	Ver <i>magnitud de la asociación</i> .

## Glosario

<b>Función de riesgo</b>	<i>Hazard function</i>	En un <i>análisis de supervivencia</i> , función que representa la probabilidad condicionada por unidad de tiempo que tiene un sujeto de fallecer en un instante determinado si se había sobrevivido hasta el instante anterior.
<b>Función de supervivencia</b>	<i>Survival function</i>	En un <i>análisis de supervivencia</i> , función que sintetiza la probabilidad de que ocurra el suceso de interés en función del tiempo transcurrido. Frecuentemente se representa en un gráfico cuyo eje horizontal corresponde al tiempo transcurrido desde el inicio del seguimiento y cuyo eje vertical presenta la probabilidad de supervivencia, habitualmente expresada en porcentaje. Dos son los métodos habitualmente utilizados para ello: el <i>método actuarial</i> y el de <i>Kaplan-Meier</i> .
<b>Gauss, curva de</b>	<i>Gaussian distribution</i>	Ver <i>distribución normal</i> .
<b>Generalización</b>	<i>Generalizability</i>	Ver <i>validez externa</i> .
<b>Gradiente biológico</b>	<i>Dose-response effect</i>	Ver <i>gradiente dosis-respuesta</i> .
<b>Gradiente dosis-respuesta</b>	<i>Dose-response effect</i>	Criterio de <i>causalidad</i> que se refiere al aumento de la magnitud del <i>riesgo</i> o de la respuesta <i>a medida que se incrementa el grado de exposición al factor de estudio</i> (en cantidad y/o tiempo). No todas las relaciones causales tienen necesariamente una relación de este tipo.
<b>Gráfico</b>	<i>Plot</i>	Dibujo en el que se representan de manera visual las <i>distribuciones de frecuencias</i> .
<b>Gráfico de cajas</b>	<i>Box and whisker plot</i>	Síntesis semigráfica de la distribución de valores de una variable. Se dibuja un rectángulo (caja) cuyos límites son el primer y tercer cuartil, y se señala en su interior la situación de la mediana. De los límites de la caja, sobresalen unas patas ( <i>whiskers</i> ) que llegan hasta los valores máximo y mínimo, o bien hasta los percentiles 2,5 y 97,5, respectivamente.
<b>Gráfico de tallo y hojas</b>	<i>Stem and leaf plot</i>	Variante del <i>histograma</i> que, además de proporcionar una imagen visual de la <i>distribución</i> , permite reconstruir los datos originales juntando los números de los tallos (a la izquierda) con los de las hojas (a la derecha). Habitualmente, cada hoja corresponde a un caso o a un grupo de casos, y se representa por el último dígito del valor de la variable de interés.
<b>Grupo control</b>	<i>Control group</i>	Grupo de comparación de un estudio. Ver <i>controles</i> .
<b>Guía de práctica clínica (GPC)</b>	<i>Clinical Practice Guideline (CPG)</i>	Conjunto de recomendaciones diseñadas para ayudar tanto a los profesionales sanitarios como a los usuarios, a seleccionar las opciones diagnósticas y/o terapéuticas más adecuadas en el enfoque de una condición clínica específica. En su elaboración es fundamental la <i>revisión sistemática</i> de la evidencia científica.
<b>Guttman, escala de</b>	<i>Guttman's scale</i>	Escala que consiste en 4 o 5 frases que expresan una actitud favorable o desfavorable hacia un tema, de intensidad creciente, de forma que quien está de acuerdo con una de ellas lo está con las anteriores.

<b>Helsinki, declaración de</b>	<i>Helsinki declaration</i>	Documento de la Asociación Médica Mundial que recoge las recomendaciones que deben guiar a los médicos en la investigación con seres humanos, adoptada en la asamblea celebrada en Helsinki en 1964 y revisada posteriormente en diferentes ocasiones.
<b>Heterogeneidad</b>	<i>Heterogeneity</i>	Los resultados de diferentes estudios son homogéneos si sus resultados no varían entre sí más de lo que sería esperable por azar. Opuesto a <i>heterogeneidad</i> .
<b>Heteroscedasticidad</b>	<i>Heteroscedasticity</i>	Desigualdad de <i>varianzas</i> (o de <i>desviaciones estándar</i> ) en la distribución de las variables en las poblaciones de las que proceden las muestras que se comparan. Opuesto a <i>homoscedasticidad</i> (asunción frecuente en las <i>pruebas paramétricas</i> ).
<b>Hipótesis alternativa (Ha)</b>	<i>Alternative hypothesis</i>	Hipótesis que se quiere demostrar fuera de toda duda con una <i>prueba estadística</i> . Es la hipótesis que se acepta si se obtiene suficiente evidencia para rechazar la <i>hipótesis nula</i> .
<b>Hipótesis nula (Ho)</b>	<i>Null hypothesis</i>	Hipótesis que se contrasta en una <i>prueba estadística</i> y que debe asumirse como cierta hasta que no haya suficiente evidencia estadística en contra. Habitualmente corresponde a la ausencia de diferencias o asociación entre los grupos que se comparan (en una <i>prueba bilateral</i> ) o a su ausencia en una sola dirección (en una <i>prueba unilateral</i> ). Opuesta a <i>hipótesis alternativa</i> .
<b>Histograma</b>	<i>Histogram</i>	Representación gráfica de la distribución de frecuencias de una serie de valores <i>cuantitativos</i> . Los valores se colocan en el eje de abscisas. Para cada uno de los valores, o intervalo de valores, se levanta un rectángulo cuya área es proporcional a la frecuencia de dicho valor o intervalo.
<b>Homogeneidad</b>	<i>Homogeneity</i>	Los resultados de diferentes estudios son heterogéneos si sus resultados varían entre sí más de lo que sería esperable por azar. Opuesto a <i>homogeneidad</i> .
<b>Homoscedasticidad</b>	<i>Homoscedasticity</i>	Igualdad de <i>varianzas</i> (o de <i>desviaciones estándar</i> ) en la distribución de las variables en las poblaciones de las que proceden las muestras que se comparan. Es una asunción frecuente en las <i>pruebas paramétricas</i> . Opuesto a <i>heteroscedasticidad</i> .
<b>Incidencia</b>	<i>Incidence</i>	Medida de la frecuencia de aparición de una enfermedad en una población en riesgo durante un período de tiempo determinado. Ver <i>incidencia acumulada</i> y <i>densidad de incidencia</i> .
<b>Incidencia acumulada</b>	<i>Cumulated incidence</i>	Proporción de individuos de una población en riesgo que desarrolla un suceso durante un período de tiempo determinado. Estima el <i>rriesgo</i> o la <i>probabilidad</i> de que un individuo desarrolle dicho suceso.
<b>Independencia</b>	<i>Independence</i>	Dos sucesos son independientes si la ocurrencia de uno de ellos no puede predecirse de ninguna manera a partir de la ocurrencia del otro. Dos variables son independientes si la distribución de los valores de una de ellas es la misma para cualquiera de los valores de la otra. Opuesto a <i>asociación</i> .

## Glosario

<b>Inferencia</b>	<i>Inference</i>	Proceso de pasar de las observaciones a las generalizaciones. En estadística, obtención de conclusiones referidas a poblaciones a partir de los datos observados en muestras.
<b>Inferencia causal</b>	<i>Causal inference</i>	Ver <i>causalidad</i> .
<b>Informador indirecto</b>		Persona próxima a un sujeto incluido en un estudio, que se utiliza como fuente de información cuando el sujeto no puede proporcionar los datos necesarios, ya sea porque ha fallecido o porque presenta trastornos mentales, por ejemplo.
<b>Intervalo de confianza (IC)</b>	<i>Confidence interval (CI)</i>	Intervalo entre cuyos límites cabe esperar que esté situado el verdadero valor del parámetro en la población de referencia, con una confianza determinada. Habitualmente se utilizan intervalos de confianza del 95%, cuyos límites están situados 1,96 <i>errores estándar</i> por encima y por debajo de la <i>estimación puntual</i> (IC95%: $X \pm 1,96 \text{ EE}$ ).
<b>Intervención</b>	<i>Intervention, maneuver</i>	Exposición o tratamiento que actúa sobre los pacientes para producir un resultado o respuesta. Cualquier acción (terapéutica, preventiva, diagnóstica) aplicada con el fin de modificar el curso de un proceso (enfermedad, situación fisiológica, hábito, etc.).
<b>Ji al cuadrado, prueba de</b>	<i>Chi-square test</i>	<i>Prueba estadística</i> para la comparación de <i>proporciones</i> , que evalúa la <i>asociación</i> entre dos <i>variables cualitativas</i> cuyos datos se presentan en una <i>tabla de contingencia</i> . Los <i>grados de libertad</i> de la prueba se determinan multiplicando el número de filas menos una por el de columnas menos una (una tabla de $3 \times 2$ tiene $[3 - 1] \times [2 - 1] = 2$ grados de libertad). Puede aplicarse cuando no más del 20% de los valores esperados son inferiores a 5, y ninguno de ellos es inferior a 2. En caso de que no se cumplan estas condiciones, la única alternativa es agrupar categorías para aumentar el número de casos de cada casilla. En tablas de $2 \times 2$ puede utilizarse la <i>corrección por continuidad de Yates</i> o la <i>prueba exacta de Fisher</i> .
<b>Justicia</b>	<i>Justice</i>	Principio <i>bioético</i> según el cual se debe tratar a todos los seres humanos con consideración y respeto, sin establecer otras diferencias entre ellos que las que redunden en beneficio de todos, y en especial de los menos favorecidos. Para ello es necesario distribuir los beneficios y las cargas de la investigación de forma equitativa. Para cumplir este principio, se han de utilizar criterios no discriminativos para la selección de los sujetos participantes.
<b>Kaplan-Meier, método de</b>	<i>Kaplan-Meier product limit method</i>	Método de cálculo de la <i>función de supervivencia</i> que tiene en cuenta el tiempo de participación de cada sujeto, determinando las proporciones exactas de supervivencia, y no aproximaciones como ocurre con el <i>método actuarial</i> . Por ello, es especialmente apropiado cuando se estudia un número reducido de pacientes.

<b>Kappa de Cohen, índice</b>	<i>Kappa index</i>	Medida de la <i>concordancia</i> entre dos mediciones de una <i>variable dicotómica</i> (p. ej., las obtenidas por dos observadores), eliminando la fracción de la concordancia debida al azar, es decir, la que se obtendría si las dos medidas no estuvieran relacionadas. Su cálculo es sencillo: $\kappa = (P_o - P_e)/(1 - P_e)$ , donde $P_o$ es la proporción total de concordancia observada y $P_e$ la esperada por azar. Su valor máximo es 1. Puede establecerse la siguiente relación entre su valor y el grado de concordancia: < 0,20: muy débil; 0,21-0,40: débil; 0,41-0,60: moderada; 0,61-0,80: buena; y 0,81-1,00: muy buena. En <i>variables políticas</i> puede calcularse un índice kappa ponderado, en el que se asignan diferentes ponderaciones a los desacuerdos según su magnitud.
<b>Kolmogorov-Smirnov, prueba de</b>	<i>Kolmogorov-Smirnov test</i>	Prueba estadística para comprobar la <i>normalidad</i> de una variable aleatoria según las observaciones realizadas en una muestra.
<b>Kruskal-Wallis, prueba de</b>	<i>Kruskal-Wallis test</i>	Prueba no paramétrica que se utiliza para comparar más de dos grupos <i>independientes</i> en caso de <i>variables ordinales</i> o de <i>variables cuantitativas</i> cuando no se cumplen las condiciones de aplicación del <i>análisis de la varianza</i> .
<b>Lectura crítica</b>	<i>Critical appraisal</i>	Proceso de evaluación e interpretación de la evidencia aportada por la literatura científica, considerando sistemáticamente la validez del diseño del estudio, la magnitud y relevancia de los resultados y su aplicabilidad a la propia práctica.
<b>Likert, escala de</b>	<i>Likert scale</i>	Escala que consiste en una serie de afirmaciones que expresan una opinión sobre un tema. Los encuestados deben expresar su grado de acuerdo o de desacuerdo con cada una de ellas.
<b>Límite de confianza</b>	<i>Confidence limit</i>	Valores que constituyen los límites superior e inferior de un <i>intervalo de confianza</i> .
<b>Logrank, prueba de</b>	<i>Logrank test</i>	Prueba no paramétrica de comparación de dos <i>curvas de supervivencia</i> . Se basa en dividir el tiempo en intervalos en función de los diferentes tiempos de supervivencia observados, ignorando las <i>observaciones incompletas</i> . Para cada uno de los intervalos, compara los datos observados en cada uno de los $k$ grupos con los esperados si la hipótesis nula de que no existen diferencias fuera cierta, de forma similar a la $\chi^2$ al cuadrado, con $k - 1$ grados de libertad. También se denomina prueba de Mantel-Cox.
<b>Longitudinal</b>	<i>Longitudinal</i>	Diseño en que existe un lapso de tiempo entre las variables estudiadas, de forma que puede establecerse una adecuada <i>secuencia temporal</i> entre ellas, incluso aunque hayan sido medidas en un mismo momento del tiempo. Pueden ser tanto <i>descriptivos</i> como <i>analíticos</i> . Algunos autores limitan el significado de este término al seguimiento de un grupo de sujetos desde una línea basal hasta el desenlace, haciéndolo sinónimo de <i>cohorte</i> . Opuesto a <i>transversal</i> .

## Glosario

<b>Magnitud de la asociación</b>	<i>Strength of the association</i>	1. Valor de la <i>medida de asociación</i> entre el <i>factor de estudio</i> y la <i>variable de respuesta</i> , que mide la fuerza de la <i>asociación</i> entre ambos. 2. Criterio de <i>causalidad</i> que se refiere a que, cuanto mayor sea el valor de la <i>medida de la asociación</i> entre el <i>factor de estudio</i> y la <i>variable de respuesta</i> , mayor es la probabilidad de que la relación entre ellos sea causal.
<b>Magnitud del efecto</b>	<i>Effect size</i>	Término genérico que se refiere a la estimación del efecto del factor de estudio sobre la variable de respuesta observado en un estudio.
<b>Mann-Whitney, prueba de</b>	<i>Mann-Whitney test</i>	<i>Prueba no paramétrica</i> que se utiliza para comparar dos grupos <i>independientes</i> en caso de <i>variables ordinales</i> o de <i>variables cuantitativas</i> cuando no se cumplen las condiciones de aplicación de la <i>t de Student</i> .
<b>MANOVA</b>	<i>MANOVA</i>	Ver <i>análisis de la varianza multivariante</i> .
<b>Mantel-Cox, prueba de</b>	<i>Mantel-Cox logrank test</i>	Ver <i>logrank, prueba de</i> .
<b>Mantel-Haenszel, prueba de</b>	<i>Mantel-Haenszel test</i>	1. Método de cálculo de la <i>ji al cuadrado</i> en un <i>análisis estratificado</i> , que determina la existencia de asociación entre dos variables cualitativas controlando el efecto de una tercera variable cualitativa o clasificada en intervalos por la que se estratifica. 2. Procedimiento para el cálculo de una <i>odds ratio ajustada</i> por una tercera variable cualitativa en un análisis estratificado.
<b>McNemar, prueba de</b>	<i>McNemar test</i>	<i>Prueba estadística</i> para la comparación de dos <i>proporciones</i> con <i>datos apareados</i> .
<b>Media</b>	<i>Mean</i>	<i>Medida de posición</i> que corresponde a un valor promedio de una serie de observaciones. Habitualmente se refiere a la <i>media aritmética</i> .
<b>Media aritmética</b>	<i>Mean</i>	<i>Medida de posición</i> de la distribución de valores de una <i>variable continua</i> que corresponde a la suma de los valores observados en una muestra dividida por el número de observaciones.
<b>Media geométrica</b>	<i>Geometric mean</i>	<i>Medida de posición</i> de la <i>distribución</i> de valores de una <i>variable continua</i> que corresponde a la suma de los logaritmos de los valores observados en una muestra dividida por el número de observaciones, y tomando después el antilogaritmo de este valor. Puede ser más apropiada que la <i>media aritmética</i> para promediar <i>cocientes</i> , cuando la distribución tiene una gran <i>asimetría</i> o cuando se trabaja con datos tras una transformación logarítmica.
<b>Media ponderada</b>	<i>Weighted mean</i>	<i>Medida de posición</i> de la distribución de valores de una <i>variable continua</i> que se obtiene multiplicando cada valor observado por un peso, sumando esos productos y dividiendo todo ello por la suma de los pesos.
<b>Mediana</b>	<i>Median</i>	<i>Medida de posición</i> de la distribución de valores de una <i>variable cuantitativa</i> que corresponde al valor que divide a la muestra ordenada de menor a mayor (o de mayor a menor) en dos partes iguales.

<b>Medicina basada en la evidencia (MBE)</b>	<i>Evidence-based Medicine (EBM)</i>	Utilización consciente, explícita y juiciosa de la mejor evidencia científica disponible proveniente de la investigación clínica para tomar decisiones sobre el manejo individual de los pacientes. Los pasos a seguir en su aplicación son: convertir las necesidades de información en preguntas susceptibles de respuesta, localización eficiente de la mejor evidencia con la que responder, valoración crítica de la validez y utilidad de dicha evidencia, aplicación de los resultados de dicha valoración a la práctica clínica y evaluación de la actuación.
<b>Medidas de asociación</b>	<i>Association measures</i>	Medidas que estiman la magnitud de una asociación comparando las <i>medidas de frecuencia</i> de la respuesta en los grupos expuesto y no expuesto al factor de estudio. Ver <i>riesgo relativo</i> y <i>odds ratio</i> .
<b>Medidas de frecuencia</b>	<i>Frequency measures</i>	Ver <i>prevalencia</i> , <i>incidencia</i> , <i>incidencia acumulada</i> y <i>densidad de incidencia</i> .
<b>Medidas de impacto</b>	<i>Impact measures</i>	Medidas que cuantifican el efecto de un <i>factor de estudio</i> sobre la frecuencia o el <i>riesgo</i> de un resultado, en términos de su impacto potencial sobre la salud de la población. Las más utilizadas son la <i>diferencia de incidencias</i> y la <i>fracción atribuible</i> .
<b>Medidas del efecto</b>	<i>Effect measures</i>	Medidas que cuantifican el <i>efecto</i> de un <i>factor de estudio</i> sobre la frecuencia o el <i>riesgo</i> de un resultado de salud. Pueden utilizarse medidas absolutas ( <i>diferencia de incidencias</i> , <i>diferencia de riesgos</i> , <i>reducción absoluta de riesgo</i> , etc.) y relativas ( <i>riesgo relativo</i> , <i>odds ratio</i> , <i>reducción relativa de riesgo</i> , etc.).
<b>Megaensayo</b>	<i>Megatrial</i>	<i>Ensayo clínico aleatorio</i> con un gran <i>tamaño de la muestra</i> , capaz de detectar efectos de pequeña magnitud. Suelen considerarse como tales los ensayos de más de 10.000 pacientes.
<b>MeSH</b>	<i>Medical Subject Headings (MeSH)</i>	Tesoro de términos descriptores utilizado por la National Library of Medicine de Estados Unidos para indexar los artículos en <i>Medline</i> . Tiene una estructura jerárquica en árbol.
<b>Metaanálisis</b>	<i>Meta-analysis</i>	Técnica estadística para combinar los resultados de diferentes estudios (ver <i>revisión sistemática</i> ).
<b>Metaanálisis acumulativo</b>	<i>Accumulative meta-analysis</i>	<i>Metaanálisis</i> en que los estudios se van añadiendo uno a uno (p. ej., según su fecha de publicación o su calidad), sintetizando nuevamente los resultados cada vez que se añade un estudio.
<b>Metarregresión</b>	<i>Metaregression</i>	Técnica de <i>análisis multivariante</i> que se utiliza para explorar la relación entre determinadas características de los estudios incluidos en un <i>metaanálisis</i> y sus resultados.

## Glosario

<b>Método actuarial</b>	<i>Actuarial method</i>	Método de cálculo de la <i>función de supervivencia</i> que consiste en agrupar los tiempos de participación de cada sujeto en intervalos de tiempo y calcular la probabilidad de supervivencia de cada uno de ellos, asumiendo que las pérdidas se han producido aleatoriamente a lo largo del intervalo (es decir, a la mitad del mismo). El cálculo tiene en cuenta que se trata de una probabilidad condicional, es decir, que para poder sobrevivir en un determinado intervalo de tiempo, debe haberse sobrevivido a los previos. De esta forma, se obtiene una <i>función de riesgo</i> que corresponde al producto de la probabilidad de supervivencia de un intervalo y de todos los intervalos previos.
<b>Moda</b>	<i>Mode</i>	Valor observado con mayor <i>frecuencia</i> en una serie.
<b>Modelo aditivo</b>	<i>Additive model</i>	Modelo en el cual el efecto combinado de varios factores es la suma de los efectos de cada uno de ellos.
<b>Modelo de Cox</b>	<i>Cox's model</i>	Técnica de <i>análisis multivariante</i> aplicable cuando los datos se analizan mediante <i>análisis de supervivencia</i> . El modelo asume que el efecto del factor de estudio sobre la razón de riesgos en la población de estudio es multiplicativo y no cambia en el tiempo, por lo que también se denomina modelo de riesgos proporcionales.
<b>Modelo de efectos aleatorios</b>	<i>Random effects model</i>	Modelo en el que los factores que se estudian han sido elegidos de forma aleatoria por el investigador (o se supone que lo han sido), con la finalidad de realizar inferencias sobre la población de la que proceden. Opuesto a <i>modelo de efectos fijos</i> .
<b>Modelo de efectos fijos</b>	<i>Fixed effects model</i>	Modelo en el que los factores que se estudian han sido elegidos de antemano por el investigador con la finalidad de realizar inferencias sobre ellos. Opuesto a <i>modelo de efectos aleatorios</i> .
<b>Modelo de regresión</b>	<i>Regression model</i>	Modelo estadístico que sintetiza la relación de dependencia entre una <i>variable de respuesta</i> (variable dependiente) y un conjunto de variables predictoras (variables independientes). Se puede utilizar para predecir la variable de respuesta, para identificar las variables independientes asociadas con la variable de respuesta, o para estimar el efecto de una variable independiente principal sobre la variable de respuesta controlando por el resto de variables independientes (potenciales <i>factores de confusión</i> ). Los modelos más utilizados son la <i>regresión lineal múltiple</i> y la <i>regresión logística</i> .
<b>Modelo multiplicativo</b>	<i>Multiplicative model</i>	Modelo en el cual el efecto combinado de varios factores es el producto de los efectos de cada uno de ellos.
<b>Modificación de efecto</b>	<i>Effect modification</i>	Fenómeno por el que una variable modifica el <i>efecto del factor de estudio</i> sobre la <i>variable de respuesta</i> . Se evalúa comparando las <i>medidas del efecto</i> obtenidas en los estratos definidos según las diferentes categorías de dicha variable modificadora del efecto (ver <i>análisis estratificado</i> y <i>análisis multivariante</i> ).

<b>Muestra</b>	<i>Sample</i>	1. Subconjunto de una población. 2. Grupo de individuos elegidos de una población de acuerdo a un criterio pre establecido, con el fin de obtener información sobre determinadas características de la misma. 3. Grupo de personas que cumplen los criterios de selección de un estudio y son realmente incluidas en él.
<b>Muestra aleatoria</b>	<i>Random sample</i>	<i>Muestra</i> obtenida mediante un proceso de <i>muestreo probabilístico</i> en que todos los elementos de la población tienen por azar la misma probabilidad, distinta de cero, de ser seleccionados. No es sinónimo de <i>muestra representativa</i> . Un muestreo probabilístico tiende a asegurar que se selecciona una muestra representativa, especialmente si la población y la muestra son de gran tamaño.
<b>Muestra representativa</b>	<i>Representative sample</i>	<i>Muestra</i> que se asemeja razonablemente en sus características a la <i>población de estudio</i> , de forma que permite estimar sin <i>sesgos</i> los parámetros poblacionales, con márgenes de error aceptables y adecuados al <i>objetivo del estudio</i> . Ver también <i>representatividad</i> .
<b>Muestreo</b>	<i>Sampling</i>	Procedimiento de selección de un determinado número de sujetos ( <i>muestra</i> ) del conjunto de todos los sujetos de un grupo particular o universo ( <i>población</i> ). Las conclusiones basadas en esta muestra sólo son válidas para la población que ha sido muestreada ( <i>validez interna</i> ). Cualquier extrapolación a poblaciones diferentes o más amplias implica un juicio de valor que no forma parte de la <i>inferencia estadística</i> ( <i>validez externa</i> ).
<b>Muestreo aleatorio simple</b>	<i>Random sampling</i>	Técnica de <i>muestreo</i> en que cada unidad de muestreo de la población tiene la misma <i>probabilidad</i> de ser escogida y en la que cada una de las posibles muestras del mismo tamaño tiene la misma probabilidad de ser escogida. Se trata de un procedimiento equiprobabilístico. Requiere que se conozca de antemano la lista completa de todas las unidades de muestreo.
<b>Muestreo consecutivo</b>	<i>Consecutive sampling</i>	Técnica de <i>muestreo no probabilístico</i> que consiste en seleccionar a los sujetos que cumplen los criterios de selección especificados en el protocolo del estudio, a medida que son identificados. Es la técnica más utilizada en los <i>ensayos clínicos</i> .
<b>Muestreo en conglomerados</b>	<i>Cluster sampling</i>	Técnica de <i>muestreo en etapas múltiples</i> en que se incluyen todas las unidades secundarias de las unidades primarias seleccionadas previamente en la primera etapa.
<b>Muestreo en etapas múltiples</b>	<i>Multistage sampling</i>	Técnica de <i>muestreo</i> que consiste en seleccionar unidades de muestreo de una población (unidades primarias) y, en una segunda etapa, obtener una muestra de cada una de las unidades primarias (unidades secundarias). Se pueden usar el número de etapas que sean necesarias y, en cada una de ellas, un método diferente de muestreo ( <i>aleatorio simple, estratificado, sistemático</i> ). Ver también <i>muestreo en conglomerados</i> . Es útil cuando la población de referencia es muy grande y dispersa.

## Glosario

<b>Muestreo estratificado</b>	<i>Stratified sampling</i>	Técnica de <i>muestreo aleatorio</i> que intenta asegurar que la muestra presente la misma distribución que la población en relación con determinadas variables, previniendo la aparición de <i>sugestos</i> . Consiste en dividir la población en estratos en función de las categorías de las variables por las que se desea estratificar (aquellas que puedan influir sobre los resultados), y escoger una muestra aleatoria en cada estrato, manteniendo las proporciones observadas en la población de referencia ( <i>muestreo aleatorio estratificado proporcional</i> ). Requiere que se conozca la distribución en la población de la(s) variable(s) empleada(s) para la <i>estratificación</i> y que la variabilidad entre los distintos estratos sea lo más grande posible, mientras que dentro de cada estrato sea lo más pequeña posible.
<b>Muestreo no probabilístico</b>	<i>Non-probabilistic sampling</i>	Técnicas de <i>muestreo</i> en que se desconoce la <i>probabilidad</i> de cada unidad de ser incluida en la muestra. Se utilizan métodos en los que no interviene el azar. Son ejemplos el <i>muestreo consecutivo</i> , el <i>muestreo por cuotas</i> o la <i>inclusión de voluntarios</i> .
<b>Muestreo por cuotas</b>	<i>Quota sampling</i>	Técnica de <i>muestreo no probabilístico</i> en que se conoce la composición general de la población de referencia en términos de características tales como la edad, el sexo, o el nivel de educación. Se determina el número de personas requeridas, o cuotas, según dichas características, y se incluyen sujetos hasta conseguir el número deseado de personas para llenar cada una de las mismas.
<b>Muestreo probabilístico</b>	<i>Probabilistic sampling</i>	Técnicas de <i>muestreo</i> en que todos los individuos candidatos (unidades de muestreo en general) tienen una <i>probabilidad</i> conocida, distinta de cero, de ser incluidos en la muestra. Se utiliza alguna forma de selección <i>aleatoria</i> para seleccionarlos. Tienden a asegurar que se obtendrá una <i>muestra representativa</i> , especialmente si la población y la muestra son de gran tamaño.
<b>Muestreo sistemático</b>	<i>Systematic sampling</i>	Técnica de <i>muestreo</i> que se basa en aplicar una regla sistemática simple. Se calcula la constante de muestreo ( <i>k</i> ), dividiendo el tamaño de la población por el deseado para la muestra, se extrae la primera unidad al azar entre las <i>k</i> primeras unidades de muestreo y se le suma la constante sucesivamente hasta completar el tamaño de la muestra. No requiere tener de antemano la lista completa y exhaustiva de toda la población candidata. Además, cuando ésta está ordenada siguiendo una tendencia conocida (de mayor a menor, de más viejo a más joven, etc.), asegura una cobertura de unidades de todos los tipos.
<b>n</b>	<i>n</i>	Ver <i>tamaño de la muestra</i> .
<b>Newman-Keuls, método de</b>	<i>Newman-Keuls procedure</i>	Procedimiento estadístico de <i>comparaciones múltiples</i> que se utiliza en el <i>análisis de la varianza</i> .
<b>No maleficencia</b>	<i>Non-maleficence</i>	Principio <i>bioético</i> que obliga a no hacer nada malo a alguien aunque lo pida. Someter a posibles riesgos, aunque sean mínimos, a los sujetos de una investigación cuando ésta no va a tener <i>valididad científica</i> , y por tanto va a ser inútil, puede considerarse maleficiente. Por tanto, un principio ético básico es diseñar, ejecutar, analizar e interpretar correctamente cualquier investigación.

<b>No respuestas</b>	<i>Nonresponders</i>	Miembros de una muestra o población de estudio que no toman parte en el mismo por cualquier motivo.
<b>Normalidad</b>	<i>Normality</i>	1. Cuando se refiere a una <i>prueba diagnóstica</i> , intervalo de resultados fuera del cual las afecciones objetivo pasan a ser altamente predecibles. Existen otras definiciones de normalidad menos útiles, basadas en criterios estadísticos, culturales, de riesgo o terapéuticos. 2. En estadística, que sigue una distribución normal o de Gauss que tiene las siguientes propiedades: continua, simétrica y asintótica; la media aritmética, la mediana y la moda coinciden; y está completamente determinada por su media aritmética y desviación estándar (el intervalo de dos desviaciones estándar alrededor de la media incluye el 95% de las observaciones, por lo que se ha utilizado como criterio de normalidad).
<b>Número de personas necesario para producir un daño (NND)</b>	<i>Number needed to harm (NNH)</i>	<i>Medida del efecto</i> que expresa el número de personas del grupo de estudio (tratadas) por cada persona que ha sufrido un daño o efecto perjudicial adicional en relación con el grupo control. Corresponde al inverso del <i>aumento absoluto de riesgo</i> ( $NND = 1/AAR$ ).
<b>Número de personas que es necesario tratar para prevenir un evento (NNT)</b>	<i>Number needed to treat (NNT)</i>	<i>Medida del efecto</i> que expresa el esfuerzo terapéutico que hay que realizar (número de personas que es necesario tratar) para prevenir una muerte o un acontecimiento determinado adicional, en relación con la intervención administrada en el grupo control. Corresponde al inverso de la <i>reducción absoluta de riesgo</i> ( $NNT = 1/RAR$ ).
<b>Núremberg, código de</b>	<i>Nuremberg code</i>	Declaración de la comunidad científica internacional realizada en 1947 en la que por primera vez se estipularon los postulados éticos básicos de la investigación clínica.
<b>Objetivo del estudio</b>	<i>Study objective</i>	Pregunta principal que se desea contestar mediante la realización de una investigación. En un <i>estudio descriptivo</i> , su formulación debe incluir el fenómeno que se desea describir, el problema de salud y la población objeto de estudio. En un <i>estudio analítico</i> , su formulación requiere la identificación del <i>factor de estudio</i> , el factor de comparación, la <i>variable de respuesta</i> y la <i>población diana</i> .
<b>Observacional</b>	<i>Observational</i>	Diseño en que el <i>factor de estudio</i> no es controlado por los investigadores, sino que se limitan a observar, medir y analizar determinadas variables, sin ejercer un control sobre el factor de estudio. La exposición puede haber sido escogida por el propio sujeto (p. ej., fumar), venir impuesta (p. ej., raza) o haber sido decidida por un profesional sanitario dentro del proceso habitual de atención sanitaria. Los diseños observacionales <i>analíticos</i> más importantes son los <i>estudios de cohortes</i> y los <i>estudios de casos y controles</i> . Opuesto a <i>experimental</i> .
<b>Observancia</b>	<i>Compliance, adherence</i>	Ver <i>cumplimiento</i> .
<b>Odds</b>	<i>Odds</i>	Cociente o <i>razón</i> entre la probabilidad de que un suceso ocurra y la de que no ocurra ( $odds = p/1 - p$ ). Su relación con una <i>proporción</i> es la siguiente: $proporción = odds/odds + 1$ , y $odds = proporción/1 - proporción$ .

## Glosario

<b>Odds preprueba</b>	<i>Pre-test odds</i>	Número de veces que es más probable que un sujeto determinado tenga el problema de salud o enfermedad de interés respecto a que no lo tenga ( <i>odds</i> ), antes de conocer el resultado de la <i>prueba diagnóstica</i> . Puede calcularse a partir de la <i>probabilidad preprueba</i> : $odds\ preprueba = probabilidad\ preprueba / (1 - probabilidad\ preprueba)$ .
<b>Odds posprueba</b>	<i>Post-test odds</i>	Número de veces que es más probable que un sujeto determinado tenga el problema de salud o enfermedad de interés respecto a que no lo tenga ( <i>odds</i> ), después de conocer el resultado de la <i>prueba diagnóstica</i> . Corresponde al producto de la <i>odds preprueba</i> y el <i>cociente de probabilidad positivo</i> . Puede convertirse fácilmente en la <i>probabilidad posprueba</i> : $Probabilidad\ posprueba = odds\ posprueba / (odds\ posprueba + 1)$ .
<b>Odds ratio</b>	<i>Odds ratio</i>	Razón entre <i>odds</i> . Es la medida de asociación entre el factor de estudio y la variable de respuesta que se utiliza en los estudios de casos y controles, dado que en este diseño, al no existir seguimiento de los sujetos, no puede determinarse la <i>incidencia</i> de la respuesta. Se calcula dividiendo el producto del número de casos expuestos y de controles no expuestos por el producto del número de casos no expuestos y de controles expuestos ( $OR = ad/bc$ ). Es una buena estimación del <i>riesgo relativo</i> cuando la enfermedad es poco frecuente ( <i>riesgo basal bajo</i> , <10%). Se interpreta como un <i>riesgo relativo</i> . Cada vez se utiliza con mayor frecuencia también en los <i>estudios de cohortes, ensayos clínicos aleatorios y metaanálisis</i> .
<b>p</b>	<i>p</i>	Nivel de significación estadística.
<b>Parámetro</b>	<i>Parameter</i>	Valor, usualmente desconocido, que caracteriza a una población o a una variable aleatoria (p. ej., edad media de la población española). Función definida sobre los valores numéricos de características medibles de una población.
<b>Patrón de referencia</b>	<i>Gold standard</i>	Método, procedimiento o medición ampliamente aceptados como los mejores disponibles.
<b>Patrón oro</b>	<i>Gold standard</i>	Ver <i>patrón de referencia</i> .
<b>Pearson, correlación de</b>	<i>Pearson's correlation</i>	Ver <i>correlación, coeficiente de</i> .
<b>Pendiente</b>	<i>Slope</i>	Referida a una <i>regresión lineal</i> , ver <i>coeficiente de regresión</i> .
<b>Percentil</b>	<i>Percentile</i>	<i>Medida de posición</i> . Se define el percentil <i>i</i> de una serie de valores como el valor que deja por debajo al 100xi % de las observaciones, una vez ordenada la serie de menor a mayor. Así, el percentil 50 corresponde al valor que deja por debajo al 50% de las observaciones, y equivale a la <i>mediana</i> .
<b>Pérdidas de seguimiento</b>	<i>Losses of follow-up</i>	Sujetos que, habiendo iniciado el seguimiento de un estudio, lo abandonan o se pierde el contacto con ellos, de forma que no es posible determinar en ellos el valor de la <i>variable de respuesta</i> . Si los motivos por los que se han producido están relacionados con el pronóstico inicial o con algún efecto derivado del <i>factor de estudio</i> , su exclusión del análisis puede producir un <i>sesgo</i> . Ver también <i>abandonos y análisis por intención de tratar</i> .

<b>Pérdidas prealeatorización</b>	<i>Pre-assignment exclusions</i>	Ver <i>exclusiones</i> .
<b>Período de blanqueo</b>	<i>Wash-out period</i>	Ver <i>período de lavado</i> .
<b>Período de lavado</b>	<i>Wash-out period</i>	Tiempo necesario para eliminar todo <i>efecto residual</i> de un tratamiento administrado previamente, durante el cual puede administrarse un placebo o ninguna intervención. Suele realizarse antes de iniciar un <i>ensayo clínico</i> para descartar los efectos del tratamiento que está recibiendo el sujeto o en los <i>diseños cruzados</i> en los períodos entre tratamientos.
<b>Período de preinclusión</b>	<i>Run-in phase</i>	Período en que los sujetos candidatos a participar en un ensayo clínico reciben un placebo o una intervención activa, con la finalidad de determinar si cumplen con el tratamiento y las pautas prescritas, si responden al placebo, u obtener una observación basal más válida, y así decidir su idoneidad para ser finalmente incluidos.
<b>Persona-tiempo</b>	<i>Person-time</i>	Medida que combina personas y tiempo, y corresponde a la suma de las unidades de tiempo que cada uno de los participantes en un estudio han sido observados o seguidos. Se utiliza como denominador en el cálculo de la <i>densidad de incidencia</i> .
<b>Peto, método de</b>	<i>Peto's method</i>	Procedimiento utilizado en un <i>metaanálisis</i> para combinar <i>odds ratio</i> . Es un <i>modelo de efectos fijos</i> .
<b>Pictograma</b>	<i>Pictogram</i>	Representación gráfica de la distribución de frecuencias de una serie de valores. Utiliza un dibujo alusivo al carácter que se representa o a sus unidades de medida. Para representar la frecuencia de cada valor, la figura debe tener un tamaño (área) proporcional a la misma, o bien se asigna un valor a la figura, y se repite ésta tantas veces como sea necesario.
<b>Placebo</b>	<i>Placebo</i>	Sustancia que carece por sí misma de acción terapéutica, pero que puede producir algún efecto curativo en el enfermo si éste la recibe convencido de que posee realmente tal acción. En los <i>ensayos clínicos aleatorios</i> , puede utilizarse un placebo de similares características organolépticas al tratamiento de estudio, que se administra de forma encubierta al grupo control con el fin de discernir los efectos farmacológicos reales de los debidos a las expectativas asociadas al tratamiento ( <i>efecto placebo</i> ).
<b>Plausibilidad biológica</b>	<i>Biological plausibility</i>	Criterio de <i>causalidad</i> que se refiere a que los resultados obtenidos deben ser compatibles con los conocimientos existentes sobre la actividad y los efectos biológicos del <i>factor de estudio</i> y la enfermedad. Algunos autores consideran suficiente con que pueda elaborarse una hipótesis razonable sobre el mecanismo de acción, mientras que otros exigen que existan suficientes evidencias sobre el mismo.
<b>Población</b>	<i>Population</i>	1. Grupo de personas con una característica determinada o de un ámbito concreto. 2. Denominador en el cálculo de una tasa.

## Glosario

<b>Población de estudio</b>	<i>Study population</i>	Subconjunto de la <i>población diana</i> al que se tiene la intención de estudiar. Se define por los criterios de selección establecidos en el protocolo y determinadas características geográficas y temporales que la hacen accesible a los investigadores.
<b>Población diana</b>	<i>Target population</i>	Población a la que se desearía generalizar los resultados de un estudio. Es el conjunto de individuos al que hace referencia el <i>objetivo del estudio</i> .
<b>Polígono de frecuencias</b>	<i>Frequency polygon</i>	Representación gráfica de la distribución de frecuencias de una serie de valores <i>cuantitativos</i> . Los valores se colocan en el eje de abscisas. A cada uno de los valores, o intervalo de valores, se le asigna un punto en el plano, cuya ordenada es su <i>frecuencia</i> . Posteriormente, se concretan los puntos mediante segmentos de recta.
<b>Posición, medidas de</b>	<i>Position measures</i>	Valores que describen cómo se encuentra el resto de los valores de una <i>variable cuantitativa</i> respecto a ellos. Ver medidas de <i>tendencia central</i> y <i>percentiles, cuartiles y deciles</i> .
<b>Potencia estadística</b>	<i>Statistical power</i>	Probabilidad de detectar como estadísticamente significativa una diferencia entre los grupos de una determinada magnitud, cuando existe en la realidad. Probabilidad de rechazar la hipótesis nula cuando es falsa. Corresponde al complementario de la probabilidad de cometer un error tipo II ( $1 - \beta$ ). Suele aceptarse que la potencia mínima de un estudio debe ser del 80%. Por ejemplo, en los estudios sobre la eficacia de una intervención sanitaria, mide el grado de seguridad existente para evitar la conclusión falsamente negativa de que una intervención no es eficaz cuando en realidad sí lo es.
<b>Precisión</b>	<i>Precision, accuracy</i>	1. Grado en que una medida se realiza sin <i>error aleatorio</i> , es decir, que proporciona resultados idénticos o muy similares cuando se aplica sobre la misma magnitud. Sinónimo de <i>fiabilidad</i> . 2. En inferencia estadística, medida del grado con que la <i>estimación puntual</i> obtenida de una muestra puede alejarse del verdadero valor poblacional simplemente por variabilidad aleatoria. Depende de la variabilidad del parámetro que se desea estimar y del <i>tamaño de la muestra</i> (a mayor tamaño, mayor precisión). Suele medirse a través del <i>error estándar</i> o del <i>intervalo de confianza</i> .
<b>Pregunta clínica</b>	<i>Clinical question</i>	En <i>medicina basada en la evidencia</i> , pregunta susceptible de respuesta, formulada a partir de la identificación de la necesidad de información clínicamente importante para el cuidado de los propios pacientes. Requiere la identificación precisa del paciente y/o del problema de salud de interés, la intervención que se va a considerar (tratamiento, causa, factor pronóstico, etc.), la intervención con la que se va a comparar (cuando es pertinente) y la variable o variables que permitan evaluar el resultado de interés clínico.
<b>Pregunta de investigación</b>	<i>Research question</i>	Ver <i>objetivo del estudio</i> .

<b>Prevalencia</b>	<i>Prevalence</i>	Proporción de individuos de una población que tienen una enfermedad o una característica determinada en un momento del tiempo.
<b>Probabilidad</b>	<i>Probability</i>	1. Frecuencia relativa de ocurrencia de un evento en una secuencia de $n$ observaciones aleatorias, cuando $n$ tiende a infinito. 2. Medida del grado de creencia en una hipótesis o afirmación, cuyo valor puede ir desde 0 hasta 1.
<b>Probabilidad preprueba</b>	<i>Pre-test probability</i>	Probabilidad de que un sujeto determinado tenga el problema de salud o enfermedad de interés antes de conocer el resultado de la <i>prueba diagnóstica</i> .
<b>Probabilidad posprueba</b>	<i>Post-test probability</i>	Probabilidad de que un sujeto determinado tenga el problema de salud o enfermedad de interés después de conocer el resultado de la <i>prueba diagnóstica</i> . Corresponde al <i>valor predictivo</i> .
<b>Proporción</b>	<i>Proportion</i>	Cociente en que el numerador está incluido en el denominador (p. ej., proporción de varones en una población). Puede adoptar valores entre 0 y 1, aunque también suele expresarse como tanto por ciento.
<b>Prospectivo</b>	<i>Prospective</i>	Término ambiguo que suele referirse al inicio cronológico del estudio previo al desarrollo de los hechos, por lo que los datos sobre éstos se recogen a medida que ocurren. Algunos autores utilizan este término para referirse a la <i>secuencia temporal</i> de exposición a desenlace, como sinónimo de <i>cohorte</i> o de <i>longitudinal</i> . Opuesto a <i>retrospectivo</i> .
<b>Protocolo</b>	<i>Protocol</i>	De investigación, documento que incluye la razón de ser del estudio, sus objetivos, diseño y métodos, las condiciones bajo las que se realizará, así como la estrategia prevista de análisis de los datos.
<b>Prueba bilateral</b>	<i>Two-sided test</i>	Prueba estadística basada en la asunción de que los datos se distribuyen en ambas direcciones a partir de un valor central. La <i>hipótesis nula</i> que contrasta es la ausencia de diferencia o asociación entre las variables, mientras que la <i>hipótesis alternativa</i> es la existencia de dicha diferencia o asociación en cualquier dirección. Es más conservadora que la <i>prueba unilateral</i> .
<b>Prueba estadística</b>	<i>Statistical test, significance test</i>	Ver <i>contraste de hipótesis</i> .
<b>Prueba no paramétrica</b>	<i>Non-parametric test, rank statistics</i>	Prueba estadística de <i>contraste de hipótesis</i> que no realiza ninguna asunción sobre la distribución de los <i>parámetros</i> poblacionales. Suele utilizarse este término para referirse a las pruebas que se utilizan para el análisis de <i>variables ordinales</i> y de <i>variables cuantitativas</i> cuando no se cumplen las condiciones de aplicación de las pruebas paramétricas (habitualmente se trata de variables que no siguen una distribución normal y de muestras de pequeño tamaño). Se basan en el análisis de los <i>rangos</i> en lugar del de los valores propiamente dichos.

## Glosario

<b>Prueba paramétrica</b>	<i>Parametric test</i>	Prueba estadística de <i>contraste de hipótesis</i> que requiere para su uso que se cumplan determinadas asunciones sobre la distribución de <i>parámetros</i> . Se utilizan en el análisis de <i>variables cuantitativas</i> , y habitualmente requieren que sigan la <i>distribución normal</i> y que las <i>variancias</i> de los grupos que se comparan sean similares.
<b>Prueba unilateral</b>	<i>One-sided test</i>	Prueba estadística basada en la asunción de que los datos sólo tienen una posible dirección de variabilidad. Por ejemplo, al comparar dos tratamientos, la <i>hipótesis nula</i> que contrasta es que uno de ellos es inferior o igual al otro, de forma que la <i>hipótesis alternativa</i> es que es superior. El valor de <i>p</i> es la mitad del que se obtiene con una <i>prueba bilateral</i> , por lo que es menos conservadora.
<b>Punto de corte</b>	<i>Cut-off point</i>	En una <i>prueba diagnóstica</i> cuantitativa, valor que se utiliza como límite para considerar el resultado como normal o anormal. Ver también <i>curva ROC</i> .
<b>Randomización</b>	<i>Randomization</i>	Anglicismo. Ver <i>asignación aleatoria</i> .
<b>Rango</b>	<i>Range, rank</i>	1. De una serie de valores, ver <i>recorrido</i> . 2. De un valor de una muestra, número de orden que ocupa en la serie de valores de la muestra una vez ordenados.
<b>Rango intercuartílico</b>	<i>Interquartile range</i>	Ver <i>recorrido intercuartílico</i> .
<b>Razón</b>	<i>Ratio</i>	Cociente en que el numerador no está incluido en el denominador (p. ej., razón entre varones y mujeres en una población).
<b>Razón de verosimilitud</b>	<i>Likelihood ratio</i>	1. En pruebas diagnósticas, ver <i>cociente de probabilidad</i> . 2. Método de comparación de dos modelos de <i>regresión logística</i> , cuando uno de ellos es un caso especial del otro (los parámetros de un modelo son un subconjunto de los del otro). Corresponde a la diferencia entre los valores $-2LL$ de ambos modelos, donde LL es el logaritmo neperiano de la verosimilitud. Sigue una distribución de la <i>ji al cuadrado</i> , y sus <i>grados de libertad</i> corresponden a la diferencia en el número de parámetros de ambos modelos. Cuando la diferencia entre los modelos es de un solo parámetro, equivale a una prueba de significación del coeficiente de regresión.
<b>Reacción adversa</b>	<i>Adverse effect, adverse drug reaction (ADR)</i>	Reacción nociva y no intencionada que se produce por la administración de una sustancia a dosis utilizadas normalmente en el ser humano para la profilaxis, el diagnóstico o el tratamiento de enfermedades, o para la modificación de una función fisiológica. Cuando resulta de un efecto farmacológico exagerado pero previsible, se denomina tipo A. Cuando aparece a consecuencia de efectos farmacológicos imprevisibles, se denomina tipo B. Algunos autores prefieren el término efecto indeseable. Antes de la comercialización del fármaco, se emplea la denominación <i>acontecimiento adverso</i> .
<b>Reclutamiento</b>	<i>Recruitment</i>	Proceso por el que se selecciona a los sujetos candidatos para participar en un estudio.

<b>Recorrido</b>	<i>Range</i>	Diferencia entre los valores mayor y menor de una serie de observaciones.
<b>Recorrido intercuartílico</b>	<i>Interquartile range</i>	Diferencia entre el tercer y primer <i>cuartil</i> de una serie de observaciones. Contiene el 50% central de las observaciones.
<b>Reducción absoluta del riesgo (RAR)</b>	<i>Absolute Risk Reduction (ARR)</i>	Medida del efecto que se utiliza en los <i>ensayos clínicos aleatorios</i> cuando la intervención experimental reduce la mortalidad o el riesgo de desarrollar una enfermedad o acontecimiento adverso. Corresponde a la diferencia entre las incidencias o riesgos observados en el grupo control y en el de intervención ( $RAR = I_0 - I_e$ ). Ver <i>diferencia de incidencias</i> .
<b>Reducción relativa del riesgo (RRR)</b>	<i>Relative Risk Reduction (RRR)</i>	Medida del efecto que se utiliza en los <i>ensayos clínicos aleatorios</i> cuando la intervención experimental reduce la mortalidad o el riesgo de desarrollar una enfermedad o acontecimiento adverso. Expresa la reducción de riesgo obtenida en relación con el grupo control. Corresponde al cociente entre la <i>reducción absoluta de riesgo</i> y la <i>incidencia</i> o riesgo observado en el grupo control ( $RRR = RAR / I_0 = (I_0 - I_e) / I_0$ ).
<b>Regresión a la media</b>	<i>Regression to the mean</i>	Tendencia de los valores de una variable, cuando son extremos, a presentar valores más cercanos a la media cuando la variable es medida repetidas veces.
<b>Regresión, ecuación de</b>	<i>Regression equation</i>	Ecuación de la <i>recta de regresión</i> . Es del tipo $y = a + bX$ . Los parámetros $a$ y $b$ se estiman mediante el <i>método de mínimos cuadrados</i> .
<b>Regresión lineal múltiple</b>	<i>Multiple linear regression</i>	Técnica de <i>análisis multivariante</i> que se utiliza cuando la variable dependiente se ha medido en una escala cuantitativa. Ver también <i>modelo de regresión</i> .
<b>Regresión logística</b>	<i>Logistic regression</i>	Técnica de <i>análisis multivariante</i> que se utiliza cuando la variable dependiente se ha medido en una escala cualitativa (habitualmente dicotómica). Ver también <i>modelo de regresión</i> .
<b>Regresión, recta de</b>	<i>Regression line</i>	Recta que mejor se ajusta a la nube de puntos que representa la relación entre dos <i>variables cuantitativas</i> , cuando se asume una relación de dependencia entre ellas (regresión de $Y$ sobre $X$ ). Ver <i>ecuación de regresión</i> .
<b>Relación causal</b>	<i>Causal relationship</i>	Ver <i>causalidad</i> .
<b>Relación dosis-respuesta</b>	<i>Dose-response relationship</i>	Ver <i>gradiente dosis-respuesta</i> .
<b>Relación temporal</b>	<i>Temporality</i>	Ver <i>secuencia temporal</i> .
<b>Relevancia clínica</b>	<i>Clinical relevance</i>	Valoración mediante el juicio clínico de la <i>magnitud del efecto</i> o la asociación observados. Suele basarse en medidas epidemiológicas del tipo de la <i>reducción absoluta de riesgo</i> o el <i>número de personas que es necesario tratar para prevenir un evento</i> . La <i>significación estadística</i> no implica necesariamente que el resultado observado sea clínicamente relevante.

## Glosario

<b>Repetibilidad</b>	<i>Repeatability</i>	Grado en que un instrumento de medida produce resultados idénticos o muy similares al aplicarse sobre la misma magnitud. Es un componente de la <i>fiabilidad</i> de una medida.
<b>Representatividad</b>	<i>Representativeness</i>	Grado en que una muestra se asemeja a la población de la que ha sido extraída. Ver <i>muestra representativa</i> .
<b>Reproducibilidad</b>	<i>Reproducibility</i>	Ver <i>repetibilidad</i> .
<b>Respeto por las personas</b>	<i>Respect for autonomy</i>	Principio <i>bioético</i> que obliga a considerar a las personas como entes con <i>autonomía</i> y a proteger a los que la tengan disminuida. El <i>consentimiento informado</i> es el procedimiento formal para aplicar este principio
<b>Restricción</b>	<i>Restriction</i>	Procedimiento utilizado para prevenir el efecto de <i>factores de confusión</i> , que consiste en restringir la inclusión a los sujetos que presentan determinados valores del potencial factor de inclusión (p. ej., sólo varones, o sólo fumadores).
<b>Retirada</b>	<i>Withdrawal</i>	Sujeto incluido en un estudio pero que es excluido del análisis de forma deliberada porque han aparecido situaciones que encierran un peligro injustificado para el mismo u otras que así lo aconsejen (reacciones adversas, inclusión por error, violaciones graves del protocolo, etc.).
<b>Retrospectivo</b>	<i>Retrospective</i>	Término que se refiere al inicio cronológico del estudio con posterioridad al desarrollo de los hechos, de forma que los datos se obtienen a partir de registros o de lo que los sujetos refieren. Opuesto a <i>prospectivo</i> .
<b>Revisión</b>	<i>Review, overview</i>	Cualquier resumen de la literatura.
<b>Revisión sistemática</b>	<i>Systematic review</i>	Revisión de la literatura en que la evidencia disponible sobre un tema ha sido sistemáticamente identificada, críticamente evaluada y sintetizada de acuerdo con unos criterios estrictos predeterminados.
<b>Riesgo</b>	<i>Risk</i>	Probabilidad de ocurrencia de un determinado suceso. Se estima habitualmente por la proporción de sujetos que desarrollan dicho suceso (ver <i>incidencia</i> ).
<b>Riesgo absoluto (RA)</b>	<i>Absolute Risk (AR)</i>	Estrictamente se refiere al riesgo de desarrollar un suceso en la población de estudio, pero habitualmente se utiliza como sinónimo de <i>diferencia de incidencias</i> .
<b>Riesgo atribuible</b>	<i>Attributable Risk</i>	Ver <i>Fracción atribuible</i> .
<b>Riesgo basal</b>	<i>Basal risk</i>	Riesgo de desarrollar una enfermedad o un acontecimiento adverso que tienen los sujetos de unas determinadas características si no se aplica una determinada intervención.
<b>Riesgo relativo</b>	<i>Relative Risk</i>	Razón o cociente entre la <i>incidencia</i> del grupo expuesto y la del no expuesto ( $RR = I_e/I_0$ ). También se denomina razón de incidencias o razón de riesgos. Indica el número de veces que es más probable que se desarrolle la respuesta o efecto en el grupo expuesto en relación con el no expuesto. Cuando es mayor de 1 indica un aumento del <i>riesgo</i> (o probabilidad) de desarrollar el suceso, y cuando es menor de 1, indica una disminución. Un valor igual a 1 indica que no hay asociación entre el factor de estudio y la respuesta.

<b>Scheffé, método de</b>	<i>Scheffe's procedure</i>	Procedimiento estadístico de <i>comparaciones múltiples</i> que se utiliza en el <i>análisis de la varianza</i> .
<b>Secuencia temporal</b>	<i>Temporality</i>	Criterio de <i>causalidad</i> que se refiere a que la exposición al <i>factor de estudio</i> ha de preceder al efecto o la respuesta por un período de tiempo compatible con los conocimientos biológicos que se tienen de la historia natural de la enfermedad.
<b>Seguridad</b>	<i>Safety</i>	Grado en que una intervención está exenta de efectos indeseables o <i>reacciones adversas</i> .
<b>Sensibilidad</b>	<i>Sensitivity</i>	Proporción de individuos enfermos que son correctamente clasificados como tales por una <i>prueba diagnóstica</i> (resultado de la prueba positivo). Una prueba muy sensible da pocos resultados falsos negativos.
<b>Sensibilidad al cambio</b>	<i>Responsiveness</i>	Cualidad de una variable o instrumento de medida para detectar cambios en el fenómeno que se desea estudiar.
<b>Serie de casos</b>	<i>Case series</i>	Estudio descriptivo de las características de un número limitado de pacientes con una determinada enfermedad o de los resultados observados en un grupo de sujetos que han recibido una determinada intervención.
<b>Sesgo</b>	<i>Bias</i>	1. Error en el diseño de un estudio, ya sea en la selección de los sujetos ( <i>sesgo de selección</i> ) o en la medición de variables ( <i>sesgo de información</i> ), que conduce a una estimación incorrecta o no válida del efecto o parámetro que se estudia. 2. De una distribución, ver <i>asimetría (skewness)</i> .
<b>Sesgo de Berkson</b>	<i>Berkson bias</i>	Forma de <i>sesgo de selección</i> que puede ocurrir en los <i>estudios de casos y controles hospitalarios</i> , cuando la combinación de exposición y enfermedad aumenta la posibilidad de ingreso hospitalario, lo que conduce a una exposición sistemáticamente más elevada en los casos hospitalarios que en los controles, distorsionando la estimación de la <i>odds ratio</i> .
<b>Sesgo de detección</b>	<i>Detection bias, ascertainment bias</i>	<i>Sesgo</i> debido a la existencia de diferencias sistemáticas entre los grupos que se comparan en la forma en que se han establecido, diagnosticado o verificado las respuestas.
<b>Sesgo de duración de la enfermedad</b>	<i>Length time bias</i>	<i>Sesgo</i> que se produce cuando existe una mayor proporción de casos de larga duración en un grupo que en el otro. En los estudios que evalúan la eficacia de una medida preventiva, puede derivar de la mayor probabilidad de detectar los casos de progresión más lenta y con un período presintomático más largo, probablemente con un mejor pronóstico. La realización de un <i>ensayo clínico aleatorio</i> evita este sesgo. En un <i>estudio de casos y controles</i> , puede ocurrir cuando se incluyen casos prevalentes (mayor supervivencia) en lugar de incidentes.
<b>Sesgo de incorporación</b>	<i>Incorporation bias</i>	<i>Sesgo</i> que se produce en los estudios que evalúan una <i>prueba diagnóstica</i> cuando la prueba que se evalúa forma parte del índice compuesto que se utiliza como <i>patrón de referencia</i> , ya que se aumentaría artificialmente su <i>validez</i> .

## Glosario

<b>Sesgo de información</b>	<i>Measurement bias o medida</i>	Sesgo que se produce cuando las mediciones de las variables de estudio son de mala calidad o cuando son sistemáticamente desiguales entre los grupos de sujetos que se comparan.
<b>Sesgo del observador</b>	<i>Observer bias</i>	Sesgo de información que se produce por la existencia de diferencias sistemáticas entre los grupos en la forma en que el observador recoge la información. En un <i>ensayo clínico</i> puede ocurrir cuando el investigador conoce el tratamiento que recibe cada sujeto ( <i>ensayo clínico abierto</i> ).
<b>Sesgo del patrón de referencia imperfecto</b>	<i>Imperfect gold standard bias</i>	Sesgo que se produce en los estudios que evalúan una <i>prueba diagnóstica</i> cuando se utiliza un <i>patrón de referencia</i> inadecuado. Si la nueva prueba fuera mejor, podría llegar a la conclusión errónea de que no es útil porque sus resultados no coinciden con los del patrón de referencia utilizado.
<b>Sesgo de memoria</b>	<i>Memory bias</i>	Sesgo de información que puede aparecer en los <i>estudios de casos y controles</i> cuando existe un diferente recuerdo de la exposición en los casos que en los controles.
<b>Sesgo de participación</b>	<i>Participation bias</i>	Sesgo que se produce en los estudios que evalúan la eficacia de una medida preventiva cuando el grupo de estudio está formado por los sujetos que han aceptado y recibido la intervención, debido a la autoselección que se ha producido. Este sesgo puede evitarse mediante la <i>asignación aleatoria</i> de los sujetos a los grupos.
<b>Sesgo de publicación</b>	<i>Publication bias</i>	Sesgo que puede cometerse en una <i>revisión</i> de la literatura al asumir que los estudios publicados son todos los realizados. No todos los estudios finalizados son publicados. De hecho, la probabilidad de que un estudio aparezca publicado es mayor en los que han observado resultados estadísticamente significativos, en los multicéntricos y en los que han tenido una fuente de financiación externa, por citar algunos de los factores más importantes. Por ello, los artículos publicados suelen ser una muestra sesgada de la investigación realizada sobre un tema.
<b>Sesgo de selección</b>	<i>Selection bias</i>	Sesgo que se produce al elegir una <i>muestra</i> que no representa de forma adecuada a la población de estudio o bien al constituir los grupos de estudio de forma que éstos no son comparables.
<b>Sesgo de sospecha diagnóstica</b>		Sesgo que se produce en los estudios que evalúan una <i>prueba diagnóstica</i> cuando la aplicación de la nueva prueba y del <i>patrón de referencia</i> no se realiza de forma independiente y <i>enmascarada</i> , de forma que los investigadores conocen los resultados obtenidos con uno de los métodos al aplicar el otro.
<b>Sesgo de verificación diagnóstica</b>	<i>Work-up bias</i>	Sesgo que puede aparecer en los estudios que evalúan una <i>prueba diagnóstica</i> cuando no se obtiene la verificación diagnóstica en todos los casos, debido a que no se mide el <i>patrón de referencia</i> en todos los sujetos, ya que la decisión de aplicarlo está condicionada por el resultado previo de la prueba que se evalúa. Conduce a una inadecuada valoración de las características operacionales de la nueva prueba.

<b>Sesgo por adelanto en el diagnóstico</b>	<i>Lead time bias</i>	Sesgo que se produce cuando el seguimiento de los grupos que se comparan no empieza en tiempos comparables. Al evaluar la eficacia de una medida preventiva, puede ocurrir si no se tiene en cuenta la existencia de un período de detección durante el cual el problema de salud puede ser detectado precozmente (adelanto del diagnóstico), de forma que uno de los grupos ha sido diagnosticado en una fase más temprana de la historia natural de la enfermedad que el otro grupo, lo que conduce a una sobreestimación del tiempo de supervivencia. La realización de un ensayo clínico aleatorio evita este sesgo.
<b>Significación estadística</b>	<i>Statistical significance</i>	Probabilidad de encontrar un resultado (efecto o asociación) igual o mayor que el observado en un estudio en el supuesto de que la hipótesis nula (habitualmente ausencia de efecto o asociación) sea cierta, determinada mediante una prueba de <i>contraste de hipótesis</i> . El <i>nivel de significación</i> estadística se denomina <i>p</i> , y habitualmente se establece en el valor 0,05. Por tanto, se considera que un resultado es estadísticamente significativo cuando el valor <i>p</i> es inferior a 0,05. La significación estadística depende tanto de la magnitud de la diferencia observada y de su variabilidad, como del <i>tamaño de la muestra</i> . Así, en estudios de gran tamaño, pequeñas diferencias pueden ser estadísticamente significativas, mientras que en estudios de pequeño tamaño, grandes diferencias pueden no serlo.
<b>Simple ciego</b>	<i>Single blind</i>	Procedimiento de <i>enmascaramiento</i> utilizado habitualmente en los <i>ensayos clínicos aleatorios</i> , de manera que una de las partes (generalmente los pacientes) desconoce o se le oculta información sobre qué tratamiento recibirá durante el estudio.
<b>Simulación</b>	<i>Sham, simulation</i>	1. Técnica utilizada en los <i>ensayos clínicos</i> para poder realizar un <i>enmascaramiento</i> de las intervenciones. Se aplica especialmente cuando se evalúan intervenciones no farmacológicas (p. ej., realizar una incisión cutánea para enmascarar una intervención quirúrgica). 2. Uso de un modelo matemático o animal como aproximación a un sistema real.
<b>Snedecor, prueba de la F de</b>	<i>Snedecor's F test</i>	Ver <i>F de Snedecor, prueba de la</i> .
<b>SnNout</b>	<i>SnNout</i>	Regla nemotécnica que se refiere a que, cuando una <i>prueba diagnóstica</i> tiene una alta <i>sensibilidad</i> , un resultado negativo hace imposible ( <i>out</i> ) el diagnóstico.
<b>Sobreemparejamiento</b>	<i>Overmatching</i>	Fenómeno que puede ocurrir al realizar un <i>emparejamiento</i> entre casos y controles, por el cual se pierde eficiencia y se disminuye la validez de la comparación. Puede ocurrir cuando se empareja por variables que no son potenciales <i>factores de confusión</i> , o por variables que son <i>intermedias</i> en la cadena causal entre el factor de estudio o las variables de respuesta, o cuando no se aplica una técnica de análisis para <i>datos apareados</i> .

## Glosario

<b>Spearman, correlación de</b>	<i>Spearman's correlation, Spearman rank correlation, Spearman's rho</i>	Prueba no paramétrica que se utiliza para determinar el grado de asociación o correlación entre dos <i>variables ordinales</i> o entre dos <i>variables cuantitativas</i> cuando no se cumplen las condiciones de aplicación de la <i>correlación lineal de Pearson</i> .
<b>SpPin</b>	<i>SpPin</i>	Regla nemotécnica que se refiere a que, cuando una <i>prueba diagnóstica</i> tiene una alta <i>especificidad</i> , un resultado positivo hace realmente posible ( <i>in</i> ) el diagnóstico.
<b>t de Student, prueba de</b>	<i>Student's t test, t-test</i>	1. Prueba estadística para la comparación de dos <i>medias</i> de variables que siguen la <i>distribución normal</i> y con <i>varianzas iguales</i> (homoscedasticidad), en el caso de muestras <i>independientes</i> . 2. Prueba estadística para la comparación de dos <i>medias</i> de variables que siguen la <i>distribución normal</i> , en el caso de muestras <i>apareadas</i> .
<b>Tabla de contingencia</b>	<i>Contingency table</i>	Presentación en forma de tabla de los valores de dos variables. Las categorías de una de ellas se indican en las filas y las de la otra en las columnas. La intersección entre cada fila y columna se denomina casilla o celda, y en su interior suele indicarse la <i>frecuencia absoluta</i> o relativa de observaciones que han presentado la combinación de las categorías de ambas variables correspondientes a la fila y columna correspondiente. La suma de los valores de cada fila y columna se denominan marginales.
<b>Tabla de vida</b>	<i>Life table</i>	Ver <i>método actuarial</i> .
<b>Tamaño de la muestra</b>	<i>Sample size</i>	Número de sujetos incluidos en un estudio. Antes de iniciar una investigación, debe determinarse el número de sujetos necesarios para tener garantías de que puede detectarse el efecto o la asociación de interés bajo unas asunciones determinadas. En un <i>estudio descriptivo</i> , se calcula en función de la variabilidad del parámetro que se desea estimar y de la <i>precisión</i> y confianza con que se desea realizar la <i>estimación</i> (ver <i>intervalo de confianza</i> ). En un <i>estudio analítico</i> , se calcula en función de la variabilidad de la <i>variable de respuesta</i> en el <i>grupo control</i> , de la mínima diferencia de relevancia clínica que se desea ser capaz de detectar, y del <i>nivel de significación</i> y <i>potencia estadística</i> deseados. En todos los casos, el resultado del cálculo debe ser corregido según el porcentaje de <i>no respuestas</i> y <i>pérdidas</i> esperables.
<b>Tasa</b>	<i>Rate</i>	Proporción que tiene en cuenta el tiempo. Suele utilizarse para medir la frecuencia de aparición de un suceso en una población por unidad de tiempo.
<b>Temporalidad</b>	<i>Temporality</i>	Ver <i>secuencia temporal</i> .
<b>Tendencia central, medidas de</b>	<i>Central tendency measures</i>	Valores que describen el centro de la <i>distribución</i> de una serie de valores de una <i>variable cuantitativa</i> . Ver <i>media, mediana y moda</i> .
<b>Teorema de Bayes</b>	<i>Bayes rule</i>	Ver <i>Bayes, teorema de</i> .
<b>Teorema del límite central</b>	<i>Central limit theorem</i>	Ver <i>límite central, teorema del</i> .
<b>Test-retest</b>	<i>Test-retest</i>	Ver <i>repetibilidad</i> .

<b>Tipificación</b>	<i>Normalization</i>	En estadística, procedimiento por el que se convierte una variable aleatoria de <i>distribución normal</i> en otra de la misma distribución, pero de <i>media = 0</i> y <i>desviación estándar = 1</i> (conocida como distribución Z o distribución normal típica). Consiste en restar la media y dividir por la desviación estándar, con lo que se obtiene un valor Z.
<b>Transversal</b>	<i>Cross-sectional</i>	Diseño en que los datos de cada sujeto representan esencialmente un momento del tiempo. Los datos pueden corresponder a la presencia, ausencia o diferentes grados de una característica o enfermedad ( <i>estudios de prevalencia</i> ) o examinar la relación entre diferentes variables en una población definida en un momento del tiempo determinado ( <i>estudios de asociación cruzada</i> ). La limitación de estos diseños es que no pueden establecer una <i>secuencia temporal</i> adecuada, por lo que no permiten abordar la evaluación de una <i>relación causal</i> . Opuesto a <i>longitudinal</i> .
<b>Tratamiento de rescate</b>	<i>Rescue treatment</i>	Terapéutica prevista en el protocolo de un <i>ensayo clínico</i> para administrar a los sujetos que no presenten un resultado adecuado con la intervención que reciben.
<b>Triple ciego</b>	<i>Triple blind</i>	Procedimiento de <i>enmascaramiento</i> utilizado habitualmente en los <i>ensayos clínicos aleatorios</i> , de manera que ni los pacientes, ni los investigadores o el personal clínico en contacto con los pacientes, ni los responsables de la monitorización o del análisis estadístico conocen a qué grupo han sido asignados los pacientes.
<b>Tukey, prueba de</b>	<i>Tukey's procedure</i>	Procedimiento estadístico de <i>comparaciones múltiples</i> que se utiliza en el <i>análisis de la varianza</i> cuando los tamaños de las muestras que se comparan son iguales.
<b>U de Mann-Whitney, prueba de la</b>	<i>Mann-Whitney U test</i>	Ver <i>Mann-Whitney, prueba de</i> .
<b>Utilidad</b>	<i>Utility</i>	En economía, satisfacción derivada de haber obtenido una determinada cantidad de un bien o artículo. Supone tener en cuenta, no sólo los resultados obtenidos con una intervención por sí mismos, sino también su valor social relativo y las preferencias personales (ver <i>análisis coste-utilidad</i> ).
<b>Validación</b>	<i>Validation</i>	Proceso para determinar el grado de <i>validez</i> de un método o instrumento de medida. Habitualmente se aplica a la evaluación de la <i>fiabilidad</i> y <i>validez</i> de un <i>cuestionario</i> .
<b>Validez</b>	<i>Validity</i>	1. Grado de solidez o rigor de un estudio, en el sentido de que el modo en que ha sido diseñado y realizado hace que sus resultados no estén sesgados y proporcionen una respuesta correcta a la pregunta planteada en su objetivo (ver <i>validez interna</i> y <i>validez externa</i> ). 2. Referida a una prueba diagnóstica, grado en que sus resultados se aproximan al verdadero diagnóstico (según el <i>patrón de referencia</i> ). 3. Grado en que una variable mide aquello para lo que está destinada (ausencia de <i>sesgo</i> ). Sinónimo de <i>exactitud</i> . 4. Referida a un cuestionario o instrumento de medida, grado con que mide el concepto o atributo que se pretende evaluar. No existe una única forma de determinarla, y suele distinguirse entre validez de contenido, validez de criterio, validez de concepto o constructo, validez convergente y validez discriminante.

## Glosario

<b>Validez aparente</b>	<i>Face validity</i>	Ver <i>validez lógica</i> .
<b>Validez concurrente</b>	<i>Concurrent validity</i>	Forma de evaluar la <i>validez de criterio</i> de un cuestionario cuando la <i>concordancia</i> entre los resultados de éste y del <i>patrón de referencia</i> se establece en el mismo momento temporal.
<b>Validez de concepto</b>	<i>Concept validity</i>	Ver <i>validez de constructo</i> .
<b>Validez de contenido</b>	<i>Content validity</i>	Grado en que un cuestionario contempla todos los aspectos relacionados con el concepto en estudio. Se basa en el análisis lógico del concepto que se pretende medir y en la definición de las áreas o dimensiones que abarca, y se determina a priori si el cuestionario contiene ítems representativos de todas ellas y si su número es proporcional a la importancia que la teoría les concede.
<b>Validez de criterio</b>	<i>Criterion validity</i>	Grado en que los resultados de un cuestionario concuerdan con los de un criterio externo independiente, del que existe un elevado consenso y tradición entre los investigadores de que es un buen indicador del concepto que se está midiendo ( <i>patrón de referencia</i> o <i>patrón oro</i> ). Puede evaluarse como <i>validez concurrente</i> o como <i>validez predictiva</i> .
<b>Validez externa</b>	<i>External validity</i>	Grado en que los resultados de un estudio pueden ser generalizados a poblaciones más amplias o extrapolados a poblaciones distintas a la estudiada. Para que un estudio tenga validez externa, debe tener primero <i>validez interna</i> .
<b>Validez interna</b>	<i>Internal validity</i>	Grado en que los resultados de un estudio son válidos (libres de error) para la población que ha sido estudiada, y que depende del diseño del estudio. Los <i>sesgos</i> y los <i>factores de confusión</i> son amenazas a la validez interna. Ver también <i>validez externa</i> .
<b>Validez lógica</b>	<i>Face validity</i>	Grado en que parece que un cuestionario, una parte de él o un ítem, mide lo que quiere medir. Es una valoración subjetiva.
<b>Validez predictiva</b>	<i>Predictive validity</i>	Forma de evaluar la <i>validez de criterio</i> de un <i>cuestionario</i> cuando el <i>patrón de referencia</i> es un acontecimiento futuro que se intenta predecir mediante los resultados del cuestionario o instrumento de medición.
<b>Valor atípico</b>	<i>Outlier</i>	Valor que difiere mucho del resto de observaciones. Puede ser debido a que exista un error en su medición o en su codificación, o a que corresponda a una observación que procede de una población distinta al resto. A veces puede dudarse si debe incluirse en el análisis, ya que tiene una gran influencia potencial sobre los resultados, pero su exclusión también limita la interpretación. Suele ser recomendable realizar el análisis con y sin estos datos, para determinar su grado de influencia sobre las conclusiones.
<b>Valor extremo</b>	<i>Outlier</i>	Ver <i>valor atípico</i> .
<b>Valor predictivo negativo</b>	<i>Negative Predictive Value</i>	Probabilidad de que un sujeto con un resultado negativo de una prueba diagnóstica no padezca realmente la enfermedad.
<b>Valor predictivo positivo</b>	<i>Positive Predictive Value</i>	Probabilidad de que un sujeto con un resultado positivo de una prueba diagnóstica padezca realmente la enfermedad.

<b>Variable</b>	<i>Variable</i>	Característica que varía dentro de un conjunto de datos.
<b>Variable aleatoria</b>	<i>Random variable</i>	Función que toma diversos valores numéricos, dependientes de los resultados de un fenómeno <i>aleatorio</i> , con distintas probabilidades.
<b>Variable categórica</b>	<i>Categorical data</i>	Variable cualitativa cuyos valores corresponden a categorías que no pueden ser ordenadas según una secuencia lógica (p. ej., raza o estado civil). Pueden clasificarse en variables dicotómicas y variables politómicas.
<b>Variable continua</b>	<i>Continuous data</i>	<i>Variable cuantitativa</i> cuyos valores pueden adoptar cualquier valor, dentro de unos límites (p. ej., peso).
<b>Variable cualitativa</b>	<i>Qualitative data</i>	Característica que no puede representarse numéricamente (p. ej., sexo, grupo sanguíneo, etc.). Pueden clasificarse en <i>variables categóricas</i> y <i>variables ordinales</i> .
<b>Variable cuantitativa</b>	<i>Quantitative data</i>	Característica cuyos valores se expresan numéricamente (p. ej., edad, peso, etc.). Puede clasificarse en <i>variable discreta</i> y <i>variable continua</i> .
<b>Variable dependiente</b>	<i>Dependent variable</i>	En un <i>análisis de regresión</i> , variable que se asume que puede predecirse (depende de) a partir de los valores de una o más <i>variables independientes</i> , predictoras o explicativas. Habitualmente se representa por Y.
<b>Variable dicotómica</b>	<i>Dichotomous data</i>	<i>Variable cualitativa</i> con sólo dos categorías o modalidades (p. ej., sexo).
<b>Variable discreta</b>	<i>Discrete data</i>	<i>Variable cuantitativa</i> cuyos valores son números aislados (p. ej., número de hijos).
<b>Variable ficticia</b>	<i>Dummy variable</i>	En un <i>análisis de regresión</i> , la inclusión de una <i>variable cualitativa</i> de k categorías como <i>variable independiente</i> se realiza como k - 1 variables ficticias <i>dicotómicas</i> , codificadas como 0 y 1. Para la categoría de referencia, todas las variables ficticias adoptan el valor 0, mientras que cada una adopta el valor 1 para cada una del resto de categorías. En el análisis de regresión, todas las variables ficticias que expresan la misma variable deben ir conjuntamente en el modelo para poder ser interpretadas.
<b>Variable independiente</b>	<i>Independent variable</i>	En un análisis de regresión, variable que se asume que puede ser utilizada para predecir o explicar los valores de otra variable con la que se supone que tiene una relación de dependencia (variable dependiente). Habitualmente se representa por X. Ver también coeficiente de regresión.
<b>Variable intermedia</b>	<i>Surrogate end-point</i>	1. Variable que se utiliza en sustitución de la <i>variable de respuesta</i> primaria de interés, porque se cree que está relacionada con ella, refleja la totalidad o gran parte del efecto del <i>factor de estudio</i> , y puede obtenerse a un coste inferior, en un período de tiempo más breve o en circunstancias en las que aquélla no podría estudiarse. Los cambios observados en la variable intermedia se supone que reflejan los que se producirían en la variable primaria. 2. Variable situada en la cadena causal entre el factor de estudio y la variable de respuesta, y por lo tanto asociada estadísticamente con ambas.
<b>Variable nominal</b>	<i>Nominal data</i>	Ver <i>variable categórica</i> .

## Glosario

<b>Variable numérica</b>	<i>Numerical data</i>	Ver <i>variable cuantitativa</i> .
<b>Variable ordinal</b>	<i>Ordinal data</i>	<i>Variable cualitativa</i> cuyas categorías son susceptibles de una ordenación lógica (p. ej., dolor leve, moderado e intenso).
<b>Variable polítómica</b>		<i>Variable cualitativa</i> con más de dos categorías (p. ej., estado civil).
<b>Variable de respuesta</b>	<i>Outcome, end-point, Response variable</i>	Variable que se utiliza para medir el efecto del factor de estudio (p. ej., mejoría, curación o prevención de una enfermedad cuando se trata de evaluar la eficacia de una intervención; o la aparición de una enfermedad o problema de salud cuando se evalúa el riesgo de una determinada exposición). Debe ser importante para el objetivo del estudio y relevante desde el punto de vista clínico, debe poderse medir con un método fiable, preciso y reproducible y en todos los sujetos de la misma manera, a la vez que lo suficientemente sensible para detectar el efecto de interés.
<b>Variable subrogada</b>	<i>Surrogate end-point</i>	Ver <i>variable intermedia</i> .
<b>Variable sustitutiva</b>	<i>Surrogate end-point</i>	Ver <i>variable intermedia</i> .
<b>Varianza</b>	<i>Variance</i>	<i>Medida de dispersión</i> que se obtiene dividiendo la suma de los cuadrados de las <i>desviaciones</i> de cada uno de los valores de una serie en relación con su <i>media</i> por el número total n de observaciones. Cuando se utiliza como estimación de la varianza de la distribución de los valores en la población, el denominador debe ser (n - 1). La raíz cuadrada de la varianza es la <i>desviación estándar</i> .
<b>Wald, prueba de</b>	<i>Wald's test</i>	Prueba de <i>contraste de hipótesis</i> de un <i>coeficiente de regresión logística</i> . La hipótesis nula es que el coeficiente es cero.
<b>Welch, prueba de</b>	<i>Welch's test</i>	Prueba estadística aproximada para la comparación de dos medias con <i>datos independientes</i> cuando las variables siguen una <i>distribución normal</i> , pero las <i>varianzas</i> son desiguales ( <i>heteroscedasticidad</i> ), por lo que no puede utilizarse la <i>t de Student</i> .
<b>Wilcoxon, prueba de</b>	<i>Wilcoxon signed-ranks test, Wilcoxon rank-sum test, Wilcoxon test for survival curves</i>	1. Prueba <i>no paramétrica</i> que se utiliza para comparar dos muestras <i>independientes</i> en caso de <i>variables ordinales</i> o de <i>variables cuantitativas</i> cuando no se cumplen las condiciones de aplicación de la <i>t de Student</i> . 2. Prueba de los rangos con signo: <i>prueba no paramétrica</i> que se utiliza para comparar dos muestras <i>apareadas</i> en caso de <i>variables ordinales</i> o de <i>variables cuantitativas</i> cuando no se cumplen las condiciones de aplicación de la <i>t de Student</i> .
<b>Yates, corrección de</b>	<i>Yates' correction</i>	Corrección <i>por continuidad</i> que se utiliza en el cálculo de la <i>ji al cuadrado</i> en tablas de <i>contingencia</i> $2 \times 2$ cuando la frecuencia esperada en una de las casillas es menor de 5 (y mayor de 2 o 3, según autores). Algunos autores aconsejan utilizarla incluso cuando todas las frecuencias esperadas son mayores de 5. Para muestras de menor tamaño, se utiliza la <i>prueba exacta de Fisher</i> .

**Z**

*Z-distribution,*  
*Z-scores, Z-test*

1. *Distribución normal tipificada*, de media 0 y desviación estándar 1. 2. Valor Z: valor *tipificado*, que resulta de restar la media de la distribución al valor observado y dividirlo por la desviación estándar. 3. *Prueba estadística* de comparación de dos proporciones, basada en la aproximación a la distribución normal de la distribución binomial. 4. Prueba estadística para la comparación de dos medias cuando el tamaño de las muestras es superior a 30 individuos (es equivalente a la *t de Student*).



## **Instrucciones para el acceso en línea**

Gracias por su compra. Este libro electrónico de Elsevier incluye el acceso a contenido *online* complementario. Por favor, haga [click aquí](#) (o vaya a [ebooks.elsevier.com](http://ebooks.elsevier.com)) para solicitar un código de activación y siga las instrucciones de registro para acceder al contenido *online*.