DISEÑO DE INVESTIGACIONES CLÍNICAS

4.ª EDICIÓN

STEPHEN B. HULLEY
STEVEN R. CUMMINGS
WARREN S. BROWNER
DEBORAH G. GRADY
THOMAS B. NEWMAN

Wolters Kluwer Health | Lippincott Williams & Wilkins

Diseño de Investigaciones Clínicas
4th Edition
Editors
Stephen B. Hulley MD, MPH

Professor and Chair, EmeritusDepartment of Epidemiology & BiostatisticsSchool of Medicine, University of California, San Francisco (California, EE.UU.)

Steven R. Cummings MD

Founding Director, San Francisco Coordinating CenterSenior Scientist, California Pacific Medical Center Research InstituteProfessor Emeritus, Department of Medicine, and of Epidemiology & BiostatisticsSchool of Medicine, University of California, San Francisco (California, EE.UU.)

Warren S. Browner MD, MPH

Chief Executive Officer, California Pacific Medical CenterAdjunct Professor, Department of Epidemiology & BiostatisticsSchool of Medicine, University of California, San Francisco (California, EE.UU.)

Deborah G. Grady MD, MPH

Professor of MedicineAssociate Dean for Clinical and Translational ResearchSchool of Medicine, University of California, San Francisco (California, EE.UU.)

Thomas B. Newman MD, MPH

Professor of Epidemiology & Biostatistics, and of PediatricsChief, Division of Clinical EpidemiologyAttending Physician, Department of PediatricsSchool of Medicine, University of California, San Francisco (California, EE.UU.)

Colaboradores
Norman Hearst MD, MPH

Professor of Family and Community MedicineSchool of Medicine, University of California, San Francisco (California, EE.UU.)Attending Physician, University of California Medical CenterSan Francisco (California, EE.UU.)

Michael A. Kohn MD, MPP

Associate Professor of Epidemiology and BiostatisticsSchool of Medicine, University of California, San Francisco (California, EE.UU.)Attending Physician, Emergency DepartmentMills-Peninsula Medical Center, Burlingame (California, EE.UU.)

Bernard Lo MD

President, The Greenwall FoundationProfessor of Medicine, EmeritusDirector of Program in Medical Ethics, EmeritusUniversity of California, San Francisco (California, EE.UU.)

Thomas Edward Novotny MD, MPH

Professor and Associate Director for Border and Global HealthGraduate School of Public HealthSan Diego State University, San Diego (California, EE.UU.)

2014
Lippincott Williams & Wilkins
530 Walnut Street, Philadelphia, PA 19106 USA
978-84-15840-86-2

Av. Carrilet, 3, 9.ª planta - Edifici D

08902 L'Hospitalet de Llobregat. Barcelona (España)

Tel.: 93 344 47 18

Fax: 93 344 47 16

e-mail: lwwespanol@wolterskluwer.com

Traducción y revisión:

M.ª Jesús del Sol Jaquotot

Licenciada en Medicina y Cirugía

Antonio Díez Herranz

Doctor en Medicina y Cirugía

Se han adoptado las medidas oportunas para confirmar la exactitud de la información presentada y describir la práctica más aceptada. No obstante, los autores, los redactores y el editor no son responsables de los errores u omisiones del texto ni de las consecuencias que se deriven de la aplicación de la información que incluye, y no dan ninguna garantía, explícita o implícita, sobre la actualidad, integridad o exactitud del contenido de la publicación. Esta publicación contiene información general relacionada con tratamientos y asistencia médica que no debería utilizarse en pacientes individuales sin antes contar con el consejo de un profesional médico, ya que los tratamientos clínicos que se describen no pueden considerarse recomendaciones absolutas y universales.

El editor ha hecho todo lo posible para confirmar y respetar la procedencia del material que se reproduce en este libro y su copyright. En caso de error u omisión, se enmendará en cuanto sea posible. Algunos fármacos y productos sanitarios que se presentan en esta publicación sólo tienen la aprobación de la Food and Drug Administration (FDA) para un uso limitado al ámbito experimental. Compete al profesional sanitario averiguar la situación de cada fármaco o producto sanitario que pretenda utilizar en su práctica clínica, por lo que aconsejamos la consulta con las autoridades sanitarias competentes.

Derecho a la propiedad intelectual (C. P. Art. 270)

Se considera delito reproducir, plagiar, distribuir o comunicar públicamente, en todo o en parte, con ánimo de lucro y en perjuicio de terceros, una obra literaria, artística o científica, o su transformación, interpretación o ejecución artística fijada en cualquier tipo de soporte o comunicada a través de cualquier medio, sin la autorización de los titulares de los correspondientes derechos de propiedad intelectual o de sus cesionarios.

Reservados todos los derechos.

Copyright de la edición en español

© 2014 Wolters Kluwer Health, S.A., Lippincott Williams & Wilkins

ISBN edición en español: 978-84-15840-86-2

Depósito legal: M-35751-2013

Edición en español de la obra original en lengua inglesa *Designing clinical research (4th ed.)* de Stephen B Hulley, publicada por Lippincott Williams & Wilkins

Copyright © 2013 Lippincott Williams & Wilkins

Two Commerce Square

2001 Market Street

Philadelphia, PA 19103

ISBN edición original: 978-1-60831-804-9

Producción: InVivo Proyectos Editoriales

Impresión: R. R. Donnelley-Shenzhen

Impreso en China

Dedication

A nuestras familias y nuestros alumnos

Introducción

Esta 4.ª edición de *Diseño de estudios clínicos* marca el 25.º aniversario de la publicación de nuestra 1.ª edición. Se ha convertido en el libro de texto de este tipo más utilizado, con más de 130 000 copias vendidas y ediciones en idiomas extranjeros publicadas en español, portugués, árabe, chino, coreano y japonés. Lo diseñamos como manual para la investigación clínica en todas sus facetas: ensayos clínicos, epidemiología observacional, ciencia aplicada, investigación orientada a patentes, ciencia conductual e investigación de servicios sanitarios. Hemos utilizado términos y principios epidemiológicos, presentado material conceptual avanzado de una forma práctica y amigable, y propuesto formas sencillas de abordar los complejos criterios que intervienen en el diseño de un estudio.

Muchos de nuestros lectores son médicos, enfermeras, farmacéuticos y otros científicos del ámbito sanitario que, durante su formación y como miembros jóvenes del cuerpo docente, están desarrollando su carrera profesional en investigación clínica y utilizan este libro como guía para diseñar y realizar sus estudios. Otros muchos son médicos en programas de residencia y estudiantes predoctorales de facultades profesionales (Medicina, Enfermería, Farmacia y Salud Pública, entre otros) que utilizan *Diseño de estudios clínicos* para poder convertirse en lectores perspicaces, capaces de aprender los puntos fuertes y las limitaciones de los estudios de investigación que dan forma a la práctica clínica de base científica. Un tercer grupo de lectores son estudiantes preuniversitarios que se preparan para solicitar su ingreso en estas escuelas y que están interesados en un futuro en el mundo de la investigación clínica.

¿Qué hay de nuevo en la 4.ª edición? La innovación más visible es el color, que, además de mejorar la estética, acelerará la comprensión de los componentes codificados en colores. Las principales revisiones de la 4.ª edición son la actualización y mejora del texto, las figuras y las tablas de todos los capítulos; los muchos nuevos ejemplos y referencias biográficas; y las nuevas secciones que abordan recientes avances en este campo. Por ejemplo:

- Se han reorganizado los capítulos sobre estudios de observación, con un capítulo entero dedicado ahora a los diversos diseños de casos y testigos, incluyendo el abordaje de densidad de la incidencia para analizar las modificaciones de los niveles de los factores de riesgo y las diferencias en el tiempo de seguimiento.

- En los capítulos sobre ensayos clínicos hay una sección ampliada sobre los ensayos de ausencia de inferioridad, que se han popularizado en la investigación de la eficacia comparativa, y se analiza con más detalle el análisis de subgrupos y la modificación del efecto.

- En el capítulo sobre el estudio de pruebas médicas hay una nueva sección sobre la creciente práctica de elaborar reglas de predicción clínica.

- En el capítulo sobre la utilización de bases de datos ya existentes se insiste en opciones atractivas para que los investigadores noveles publiquen de una manera rápida y económica.

- Se ha actualizado el capítulo sobre la ética de la investigación, para reflejar las políticas actuales sobre el secuenciado pangenómico y otros temas, con nuevos casos que ilustran la resolución de los dilemas éticos en la investigación clínica.

- El capítulo sobre gestión de los datos se ha actualizado de manera extensa con los últimos abordajes basados en Internet.

- En el capítulo sobre financiación hay estrategias para enfrentarse a los nuevos requisitos para la solicitud de subvenciones a los NIH, además de actualizaciones sobre la financiación por fundaciones y promotores corporativos.

La 4.ª edición está acompañada por una página web en inglés del libro en www.epibiostat.ucsf.edu/dcr/, que contiene materiales para enseñar el diseño de estudios clínicos, con vínculos a un programa detallado para los talleres de DEC de 4 y 7 semanas que impartimos a 300 alumnos cada año en la UCSF. También hay notas de los instructores para los talleres, que serán útiles para los docentes que impartan este material, y enlaces a nuestro programa de grado de maestría de formación en investigación clínica (*Training In Clinical Research,* TICR) en la UCSF, con más de 30 cursos adicionales y sus materiales. Además, hay herramientas útiles para los investigadores, entre ellas una excelente calculadora interactiva del tamaño de la muestra.

Hay muchas cosas que no han cambiado en la 4.ª edición. Éste sigue siendo un libro sencillo que omite los aspectos técnicos innecesarios e invita al investigador a que se centre en lo importante: cómo encontrar una buena pregunta de la investigación y planificar un diseño eficiente, eficaz y ético. Los capítulos sobre la estimación del tamaño de la muestra siguen desmitificando el proceso y permiten que lectores con una formación mínima en estadística hagan estos cálculos ellos mismos de manera meditada y sin necesidad de luchar con fórmulas. El libro sigue funcionando de manera óptima cuando se combina con el ingrediente esencial de uno o más mentores a largo

plazo. Todavía *no* aborda las importantes áreas de cómo analizar, presentar y publicar los hallazgos de la investigación clínica, temas que nuestros lectores pueden buscar en otros libros (por ejemplo, 1 , 2 , 3 and 4).

El proceso de transformarse en un científico clínico independiente puede ser difícil, especialmente conseguir una subvención importante por primera vez. Pero es gratificante que muchos de nuestros antiguos alumnos que utilizaron este libro han conseguido este objetivo, han descubierto que les *gusta* investigar, y se han orientado hacia una gran carrera profesional. Para los que tienen mentes inquisitivas, la búsqueda de la verdad puede convertirse en una fascinación que dura toda la vida. Para los perfeccionistas y los artesanos, hay retos interminables en la creación de estudios elegantes que respondan de manera concluyente a preguntas, grandes y pequeñas, con un coste asequible en cuanto a tiempo y dinero. Los investigadores que disfrutan del trabajo en equipo llegarán a tener relaciones gratificantes con los compañeros, el personal y los estudiantes, además de hacerse amigos de sus colaboradores que trabajan en el mismo campo en lugares distantes. Y para aquellos cuya ambición es hacer una contribución duradera a la sociedad, existe la perspectiva de que con habilidad y tenacidad puedan participar en los avances incrementales de la práctica clínica y de salud pública que forman parte del orden natural de nuestra ciencia.

BIBLIOGRAFÍA

1. Vittinghoff E, Glidden DV, Shiboski SC, et al. Regression methods in biostatistics: linear, logistic, survival, and repeated measures models, 2nd ed. New York: Springer-Verlag, 2011. Citado aquí

2. Katz MH. Multivariable analysis: a practical guide for clinicians and public health researchers, 3rd ed. New York: Cambridge University Press, 2011. Citado aquí

3. Newman TB, Kohn MA. Evidence-based diagnosis. Cambridge, MA: Cambridge University Press, 2009. Citado aquí

4. Browner WS. Publishing and presenting clinical research, 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2012. Citado aquí

Agradecimientos

Nuestro agradecimiento a la Andrew P. Mellon Foundation, por unirse a nosotros hace 30 años para comenzar el viaje de 5 años para el desarrollo de materiales docentes que se convirtieron en la 1.ª edición; a nuestro editor, por insistir constantemente en una 4.ª edición hasta que la resistencia se hizo inútil, y por poner a nuestra disposición a unos profesionales con un talento y un apoyo excepcionales que nos han ayudado a reunir todo el material; a nuestras familias, por su paciente apoyo cuando trabajábamos en esta obra; a muchos compañeros en la UCSF y otros lugares, cuyas ideas y facultades nos han influido; a nuestros estudiantes de los últimos años, cuyos logros hemos disfrutado, al tiempo que han estimulado nuestras ideas, y a nuestros lectores, responsables del éxito de este libro.

CAPÍTULO 1 Cómo empezar: anatomía y fisiología de las investigaciones clínicas

Stephen B. Hulley
Thomas B. Newman
Steven R. Cummings

Este capítulo presenta la investigación clínica desde dos puntos de vista, estableciendo temas que discurren a la par a lo largo del libro. Uno de ellos es la anatomía de la investigación: de qué está hecha. Incluye los elementos tangibles del plan del estudio: la pregunta tema del estudio, el diseño, los participantes, las mediciones o determinaciones, el cálculo del tamaño de la muestra, etc. El objetivo de un investigador es crear estos elementos de tal manera que el proyecto sea viable y eficiente.

El otro tema es la fisiología de la investigación: cómo funciona. Los estudios son útiles en la medida en que proporcionan inferencias o deducciones válidas, primero sobre lo que ocurrió en la muestra del estudio, y después sobre cómo se pueden generalizar estos hallazgos del estudio a las personas que no participan en él. El objetivo es reducir al mínimo los errores, aleatorios y sistemáticos, que amenazan a las conclusiones basadas en estas inferencias.

La separación de ambos temas es artificial, del mismo modo que la anatomía del cuerpo humano no tiene mucho sentido sin comprender algo su fisiología. Aunque la separación tiene la misma ventaja: aclara nuestras ideas sobre un tema complejo.

▪ ANATOMÍA DE LA INVESTIGACIÓN: DE QUÉ ESTÁ HECHA

La estructura de un proyecto de investigación empieza en su protocolo, o plan del estudio por escrito. Los protocolos son elementos que se usan para solicitar financiación y la aprobación del comité de ética de la investigación (CEI), pero también poseen una función científica esencial: ayudar al investigador a organizar su estudio de un modo lógico, dirigido y eficaz. En la tabla 1-1 se esbozan los componentes de un protocolo. Presentamos aquí todos ellos, se ampliarán todos los componentes en los siguientes capítulos del libro, y se volverán a reunir todas las piezas en el capítulo 19.

Pregunta de la investigación

La pregunta de la investigación es el objetivo del estudio, la incertidumbre que el investigador desea resolver. A menudo estas preguntas se inician con una preocupación general que debe reducirse a un tema concreto, que se pueda investigar. Considere, por ejemplo, la siguiente pregunta general:

- ¿Deben las personas comer más pescado?

Es un buen punto para empezar, pero la pregunta deberá concretarse antes de poder empezar con la planificación. Con frecuencia esto supone fragmentar la pregunta en componentes más específicos, y elegir uno o dos de ellos para elaborar el protocolo a su alrededor:

- ¿Con qué frecuencia comen pescado los estadounidenses?
- ¿El consumo de pescado reduce el riesgo de sufrir enfermedades cardiovasculares?
- ¿Existe riesgo de intoxicación por mercurio si los ancianos aumentan el consumo de pescado?
- ¿Tienen los complementos de aceite de pescado los mismos efectos sobre las enfermedades cardiovasculares que el pescado de la dieta?
- ¿Qué complementos de aceite de pescado no hacen que las personas huelan a pescado?

COMPONENTES DEL DISEÑO	OBJETIVO
Preguntas de la investigación	¿Qué preguntas abordará el estudio?
Introducción e importancia	¿Por qué son importantes estas preguntas?
Diseño	¿Cómo se estructura el estudio?

Marco temporal	
Diseño epidemiológico	
Participantes	¿Quiénes son los participantes y cómo se seleccionarán?
Criterios de selección	
Diseño del muestreo	
Variables	¿Qué mediciones se realizarán?
Variables predictivas	
Variables de confusión	
Variables de respuesta	
Datos estadísticos	¿Qué tamaño tiene el estudio y cómo se analizará?
Hipótesis	
Tamaño de la muestra	
Método analítico	

Una buena pregunta debe superar la prueba «¿Y qué?». Lograr la respuesta debe contribuir de forma provechosa a nuestros conocimientos. El acrónimo FINER indica cinco características esenciales de una buena pregunta de investigación: debe ser factible, interesante, novedosa, ética y relevante (cap. 2).

Antecedentes e importancia

En una breve sección de un protocolo sobre los antecedentes y la importancia se describe el estudio propuesto dentro de un contexto y se proporciona su fundamento: ¿Qué se conoce del tema que se aborda? ¿Por qué es importante la pregunta de la investigación? ¿Qué tipo de respuestas proporcionará el estudio? En esta sección se mencionan investigaciones anteriores importantes (incluso el trabajo de los propios investigadores), y se señalan los problemas que surgieron y qué cuestiones quedan sin resolver. Se especifica aquí también de qué forma los hallazgos del estudio propuesto ayudarán a resolver estas cuestiones, conducirán a nuevos conocimientos científicos o influirán en las directrices sobre la práctica clínica o las políticas de salud pública. Con frecuencia, la revisión de la literatura y la síntesis que se realice para la sección sobre la importancia del problema llevan al investigador a modificar la pregunta de la investigación.

Diseño

El diseño de un estudio es un tema complejo. Una decisión fundamental será si se opta por adoptar una actitud pasiva en la obtención de las mediciones de los participantes en el estudio en un estudio de observación, o por realizar una intervención y examinar sus efectos en un ensayo clínico (tabla 1-2). Entre los estudios de observación, dos diseños habituales son los estudios de cohortes, en los que las observaciones se realizan en un grupo de personas a las que se evalúa durante un período de seguimiento, y los estudios transversales, en los que las observaciones se realizan en una sola ocasión. Los estudios de cohortes pueden dividirse en estudios prospectivos, que se inician en el presente y siguen a los participantes en el futuro, y estudios retrospectivos, que examinan la información recopilada durante un período de tiempo en el pasado. Una tercera opción habitual es el diseño de casos y testigos, en el que el investigador compara un grupo de personas con una enfermedad u otro criterio de valoración con otro grupo que no lo presenta. Entre las opciones de ensayos clínicos, el ensayo aleatorizado y con enmascaramiento suele ser el mejor diseño, aunque los diseños no aleatorizados o sin enmascaramiento pueden ser lo único de lo que se disponga para algunas preguntas de investigación.

No existe un abordaje que sea siempre mejor que los otros, y cada pregunta de investigación precisa la determinación de cuál de los diseños será el modo más eficaz de lograr una respuesta satisfactoria. A menudo se sostiene que el estudio aleatorizado y con enmascaramiento es el mejor diseño para establecer la causalidad y la eficacia de las intervenciones, pero hay muchas situaciones en las que un estudio de observación es una mejor elección o la única opción viable. El coste relativamente bajo de los estudios de casos y testigos, y su adecuación para resultados poco frecuentes, hacen que sean atractivos para algunas preguntas. Para elegir diseños para el estudio de pruebas diagnósticas se aplican consideraciones especiales. Estos temas se comentan en los capítulos 7 a 12, en cada uno de los cuales se aborda un tipo concreto de diseño.

DISEÑO EPIDEMIOLÓGICO	CARACTERÍSTICA PRINCIPAL	EJEMPLO
	Diseños de observación	
Estudio de cohortes	Se identifica a un grupo de participantes al comienzo, y se los sigue a lo largo del tiempo	El investigador mide el consumo de pescado en un grupo de participantes al principio, y los revisa periódicamente en visitas de seguimiento para ver si aquellos que consumen más pescado tienen menos episodios de cardiopatía isquémica (CI).
Estudio transversal	Se estudia a un grupo en un punto temporal	El investigador entrevista a un grupo de participantes sobre el consumo de pescado en el pasado y en la actualidad, y correlaciona los resultados con los antecedentes de CI y la puntuación de calcio coronario actual.
Estudio de casos y testigos	Se seleccionan dos grupos por la presencia o ausencia de un resultado	El investigador examina a un grupo de pacientes con CI (los «casos») y los compara con un grupo que no tiene CI (los «testigos»), preguntando por el consumo de pescado en el pasado.
	Diseño del estudio clínico	
Estudio aleatorizado con enmascaramiento	Se crean dos grupos mediante un proceso aleatorio, y se aplica una intervención con enmascaramiento	El investigador asigna aleatoriamente a los participantes a recibir complementos de aceite de pescado o un placebo con un aspecto idéntico, y después sigue ambos grupos de tratamiento durante varios años para observar la incidencia de CI.

Una secuencia típica para estudiar un tema se inicia con estudios de observación de un tipo que con frecuencia se denomina descriptivo. En estos estudios se explora la «configuración del terreno», por ejemplo, describiendo distribuciones de características relacionadas con la salud en la población:

- ¿Cuál es el número medio de raciones de pescado a la semana en la dieta de los estadounidenses con antecedentes de cardiopatía isquémica (CI)?

A los estudios descriptivos los suelen seguir o acompañar estudios analíticos, que evalúan asociaciones para poder realizar inferencias sobre relaciones de causa y efecto:

- ¿Las personas con CI que consumen mucho pescado tienen menor riesgo de infarto de miocardio recurrente que las personas con antecedentes de CI que consumen pescado con poca frecuencia?

El paso final es, con frecuencia, un ensayo clínico para establecer los efectos de una intervención:

- ¿El tratamiento con cápsulas de aceite de pescado reduce la mortalidad total en las personas con CI?

Los ensayos clínicos suelen realizarse relativamente tarde en una serie de estudios de investigación sobre una determinada pregunta,

porque tienden a ser más difíciles y caros, y responden de manera más concluyente las preguntas más concretas que surgen de los hallazgos de los estudios de observación.

Resulta útil caracterizar un estudio en *una sola frase que resuma el diseño y la pregunta de la investigación*. Si el estudio tiene dos fases importantes, debe mencionarse el diseño de cada una de ellas:

- Este es un estudio transversal de los hábitos dietéticos de personas de 50 a 69 años de edad con antecedentes de CI, seguido por un estudio de cohortes prospectivo sobre si el consumo de pescado se asocia a un riesgo menor de sufrir episodios coronarios en el futuro.

Esta frase es el análogo en investigación de la frase que inicia un informe médico de un residente sobre un nuevo ingreso hospitalario: «Mujer policía caucásica de 62 años de edad que se sintió bien hasta 2 h antes de su ingreso, momento en el que presentó dolor torácico opresivo que se irradiaba al hombro izquierdo.»

Algunos diseños no encajan fácilmente en las categorías enumeradas anteriormente, y su clasificación con una sola frase puede resultar sorprendentemente difícil. Merece la pena el esfuerzo: una descripción precisa del diseño y de la pregunta de la investigación aclara las ideas del investigador, y resulta útil para orientar a los colaboradores y asesores.

Participantes del estudio

Deben tomarse dos decisiones importantes al elegir los participantes del estudio (cap. 3). La primera es especificar los criterios de inclusión y exclusión que definen la población a estudiar: los *tipos* de personas más adecuados para la pregunta de la investigación. La segunda decisión se refiere al modo de reunir un *número* adecuado de personas a partir de un subgrupo accesible de esta población para que participen en el estudio. Por ejemplo, para el estudio sobre el consumo de pescado en personas con CI, se podría identificar a pacientes atendidos en la consulta con códigos diagnósticos de infarto de miocardio, angioplastia o injerto de derivación arterial coronaria en la historia clínica electrónica. Las decisiones sobre qué pacientes se deben estudiar representan concesiones; el estudio de una muestra al azar de personas con CI de todo el país (o al menos de varios estados y entornos de atención sanitaria diferentes) aumentaría la posibilidad de generalización, pero sería mucho más difícil y costoso.

Variables

Otro grupo importante de decisiones en el diseño de cualquier estudio se refiere a la elección de las variables que se van a medir (cap. 4). Un estudio sobre el consumo de pescado en la dieta, por ejemplo, podría preguntar sobre los diferentes tipos de pescado que contienen diferentes concentraciones de ácidos grasos omega 3, e incluir preguntas sobre el tamaño de la ración, si el pescado estaba frito o asado, y sobre el consumo de complementos de aceite de pescado.

En un estudio analítico, el investigador estudia las asociaciones entre variables para predecir respuestas y extraer inferencias sobre causa y efecto. Al considerar la asociación entre dos variables, la que se observa primero o es más probable que sea causal partiendo de supuestos biológicos se denomina variable predictiva; la otra se denomina variable de respuesta [1] . La mayor parte de los estudios de observación tienen muchas variables predictivas (edad, raza, sexo, antecedentes de tabaquismo, consumo de pescado y complemento de aceite de pescado) y varias variables de respuesta (infartos, accidentes cerebrovasculares, calidad de vida, mal olor).

Los ensayos clínicos examinan los efectos de una intervención (un tipo especial de variable predictiva que el investigador manipula), como el tratamiento con cápsulas de aceite de pescado. Este diseño permite observar los efectos sobre la variable de respuesta usando la aleatorización para reducir al mínimo la influencia de las variables de confusión: otros factores que pueden influir en la respuesta, como el tabaquismo y el nivel de ingresos, que podrían estar asociados con el pescado de la dieta e introducir confusión en la interpretación de los hallazgos.

Aspectos estadísticos

Los investigadores deben elaborar planes para calcular el tamaño de la muestra y para gestionar y analizar los datos del estudio. Esto conlleva, generalmente, la especificación de una hipótesis (cap. 5):

Hipótesis: las mujeres de 50 a 69 años de edad con CI que tomen complementos de aceite de pescado presentarán un menor riesgo de sufrir infarto de miocardio recurrente que aquellas que no lo hacen.

Esta es una versión de la pregunta de la investigación que constituye la base para estudiar la significación estadística de los hallazgos. La hipótesis permite también al investigador calcular el tamaño de la muestra: cantidad de personas necesarias para observar la diferencia esperada en la respuesta entre grupos de estudio con una probabilidad razonable (atributo que se conoce como potencia) (cap. 6). En los estudios puramente descriptivos (¿qué proporción de personas con CI toma complementos de aceite de pescado?) no se realizan

pruebas de significación estadística y, por tanto, no precisan una hipótesis; en lugar de ello, puede calcularse el número de personas necesario para obtener intervalos de confianza aceptablemente estrechos para medias, proporciones u otros estadísticos descriptivos.

Volver al principio

▪ FISIOLOGÍA DE LA INVESTIGACIÓN: CÓMO FUNCIONA

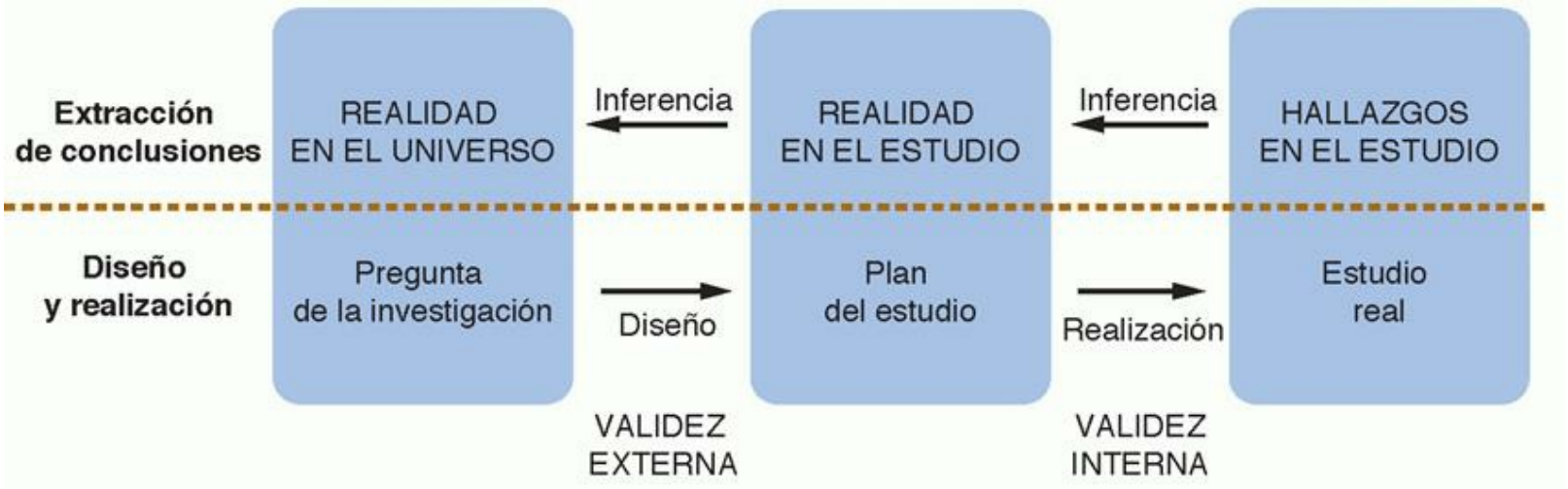
El objetivo de la investigación clínica es extraer inferencias de los hallazgos del estudio sobre la naturaleza del universo que les rodea (fig. 1-1). En la interpretación de un estudio intervienen dos grupos principales de inferencias (se ilustran de derecha a izquierda en la fig. 1-1). La inferencia n.º 1 alude a la validez interna, o grado en que el investigador extrae las conclusiones correctas sobre lo que realmente ocurrió en el estudio. La inferencia n.º 2 alude a la validez externa (también llamada posibilidad de generalización), o grado en que estas conclusiones pueden aplicarse de forma adecuada a las personas y sucesos no incluidos en el estudio.

Cuando un investigador planifica un estudio, invierte el proceso, trabajando de izquierda a derecha en la mitad inferior de la figura 1-1 , con el objetivo de aumentar al máximo la validez de estas inferencias al final del estudio. Diseña un plan de estudio en el que la elección de la pregunta de la investigación, los participantes y las mediciones aumenta la validez externa del estudio y conduce a su realización con un alto grado de validez interna. En las siguientes secciones se aborda el diseño y la realización antes de volver a los errores que amenazan la validez de estas inferencias.

Diseño del estudio

Considere esta sencilla pregunta descriptiva:

¿Cuál es la prevalencia del consumo diario de complementos de aceite de pescado en personas con CI?



▪FIGURA 1-1. El proceso del diseño y la realización de un proyecto de investigación prepara el escenario para extraer conclusiones basadas en inferencias de los hallazgos.
Herramientas de imágenes

Esta pregunta no puede contestarse con una exactitud perfecta, porque sería imposible estudiar a todos los pacientes con CI, y los métodos para descubrir si una persona tiene CI y toma aceite de pescado son imperfectos. Por lo tanto, el investigador se contenta con una pregunta relacionada que *pueda* responderse en el estudio:

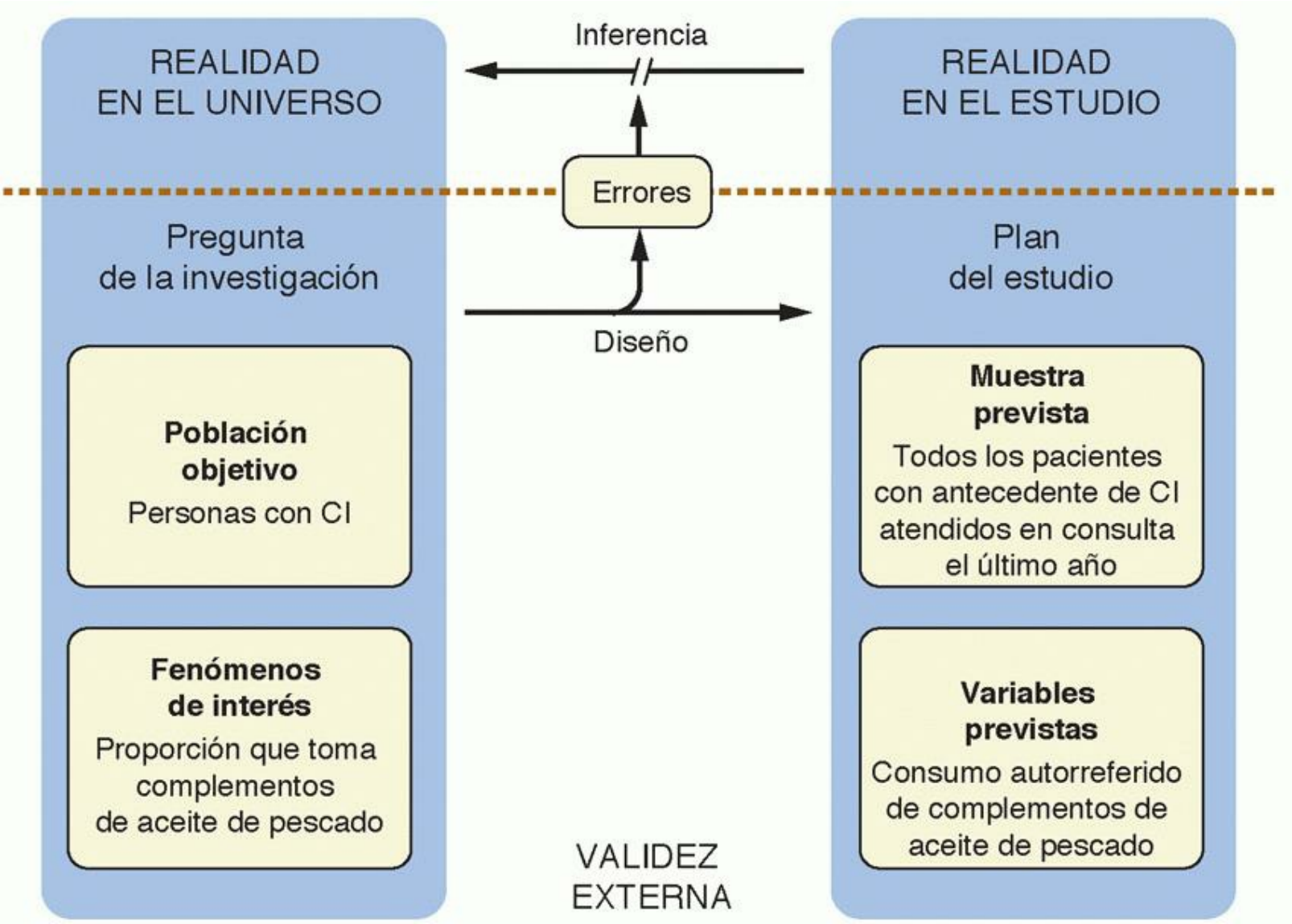
En una muestra de pacientes atendidos en la consulta del investigador que presentan un diagnóstico anterior de CI y que respondieron a un cuestionario enviado por correo, ¿qué proporción comentó que tomaba a diario complementos de aceite de pescado?

En la figura 1-2 , se ilustra la transformación desde la pregunta de investigación al plan del estudio. Un componente importante de esta transformación es la elección de una muestra de participantes que representarán a la población. El grupo de participantes especificado en el protocolo puede ser simplemente una muestra de la población de interés, porque existen barreras prácticas para estudiar a toda la población. La decisión de estudiar pacientes de la consulta del investigador identificados a través del sistema de historias clínicas electrónicas es una concesión. Es una muestra que puede estudiarse, pero tiene la desventaja de que puede proporcionar una frecuencia de consumo de aceite de pescado diferente de la que se encuentra en todas las personas con CI.

El otro componente importante de la transformación es la elección de las variables que representarán los fenómenos de interés. Las

variables especificadas en el plan del estudio suelen ser marcadores indirectos de estos fenómenos. La decisión de usar un cuestionario de autoevaluación para valorar el consumo de aceite de pescado es una forma rápida y barata de recopilar información, pero es improbable que sea perfectamente exacto, porque habitualmente las personas no recuerdan con exactitud ni registran cuánto toman en una semana típica.

En resumen, cada una de las diferencias de la figura 1-2 entre la pregunta de investigación y el plan del estudio tiene como finalidad hacer que el estudio sea más práctico. El coste de este aumento de la utilidad práctica, sin embargo, es el riesgo de que los cambios del diseño puedan hacer que el estudio lleve a una conclusión errónea o engañosa, porque está diseñado para responder a una pregunta algo diferente a la pregunta de interés de la investigación.



▪FIGURA 1-2. Errores de diseño y validez externa: si la muestra y las variables previstas no representan lo suficiente la población objetivo y los fenómenos de interés, estos errores pueden alterar las inferencias sobre lo que realmente ocurre en la población.
Herramientas de imágenes

Realización del estudio

Retomando la figura 1-1 , el lado derecho se dedica a la realización y el grado en que el estudio real cumple el plan del estudio. La cuestión aquí es el problema de una respuesta incorrecta a la pregunta de la investigación, porque el modo en que se consiguió la muestra y se realizaron las medidas difería de forma notable del modo en que estaban diseñadas (fig. 1-3).

La muestra real de participantes del estudio casi siempre es diferente de la muestra que se pretende. Los planes para estudiar, por ejemplo, a todos los pacientes de la consulta con CI elegibles podrían alterarse por diagnósticos incompletos en la historia clínica electrónica, direcciones erróneas del cuestionario enviado por correo o rechazo a participar. Las personas que aceptan participar pueden tener una frecuencia de consumo de aceite de pescado diferente de la que tienen las que no están interesadas en hacerlo. Además de estos problemas con los participantes, las mediciones reales pueden diferir de las pretendidas. Si el formato del cuestionario no está claro, las personas pueden confundirse y marcar la casilla equivocada o, simplemente, pueden omitir la pregunta por error.

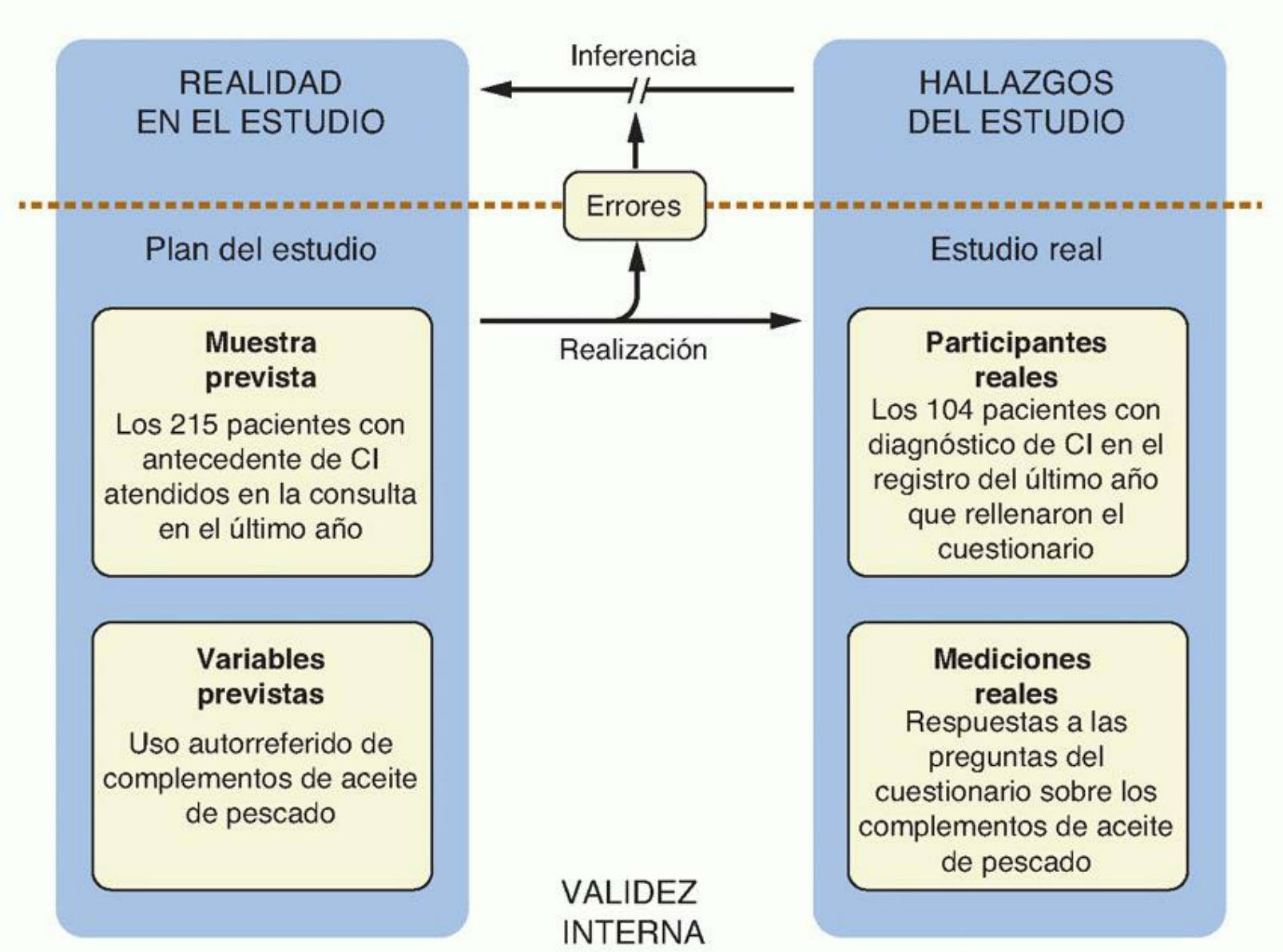
Estas diferencias entre el plan del estudio y el estudio real pueden alterar la respuesta a la pregunta de la investigación. La figura 1-3 muestra que los errores de la ejecución del estudio se juntan a errores del diseño en el camino hacia una respuesta errónea o engañosa a la pregunta de la investigación.

Inferencia causal

Un tipo especial de problema de validez surge en los estudios que examinan la asociación entre una variable predictiva y una variable de respuesta con el fin de obtener una inferencia causal. Si un estudio de cohortes halla una asociación entre el consumo de pescado y episodios de CI, ¿representa esto una relación de causa y efecto, o es el pescado un inocente espectador en una red de causalidad en la que intervienen otras variables? La reducción de la probabilidad de confusión y de otras explicaciones alternativas es uno de los principales retos que tiene el diseño de un estudio de observación (cap. 9).

Errores de la investigación

Tras reconocer que ningún estudio está libre de errores, el objetivo es aumentar al máximo la validez de las inferencias extraídas de lo que se observó en la muestra del estudio a la naturaleza de lo que ocurre en la población. Las inferencias erróneas pueden abordarse en la fase del análisis de la investigación, pero una mejor estrategia será centrarse en el diseño y la ejecución (fig. 1-4), evitando, desde el principio, que se produzcan errores en la medida que esto sea práctico.



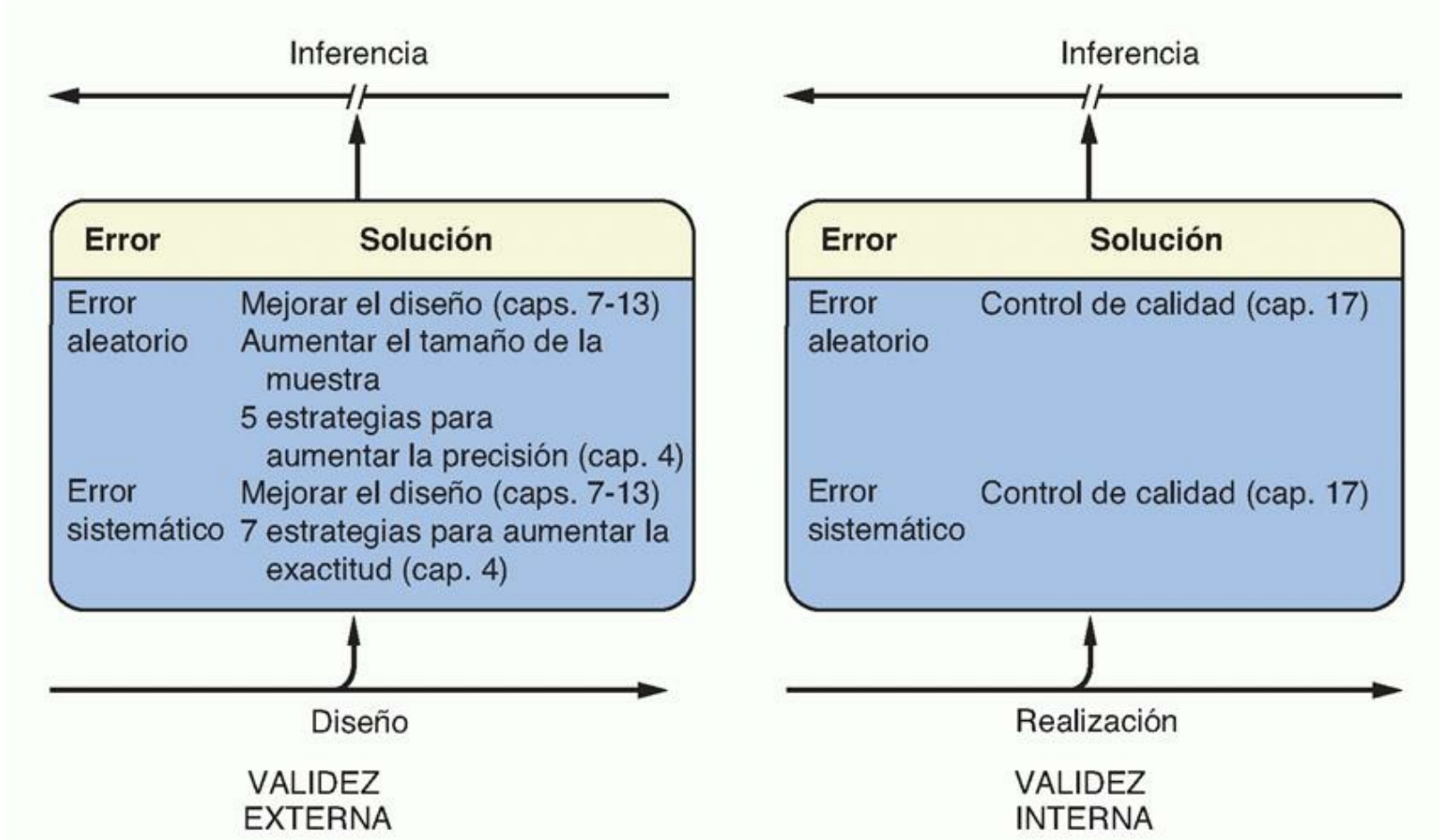
▪FIGURA 1-3. Errores de realización y validez interna: si los participantes y las mediciones reales no representan lo suficiente la muestra y las variables previstas, estos errores pueden alterar las inferencias sobre lo que ha ocurrido en el estudio.
Herramientas de imágenes

Los dos tipos principales de errores que interfieren con las inferencias de la investigación son el error aleatorio y el error sistemático. La distinción es importante, porque las estrategias para reducirlos al mínimo son bastante diferentes.

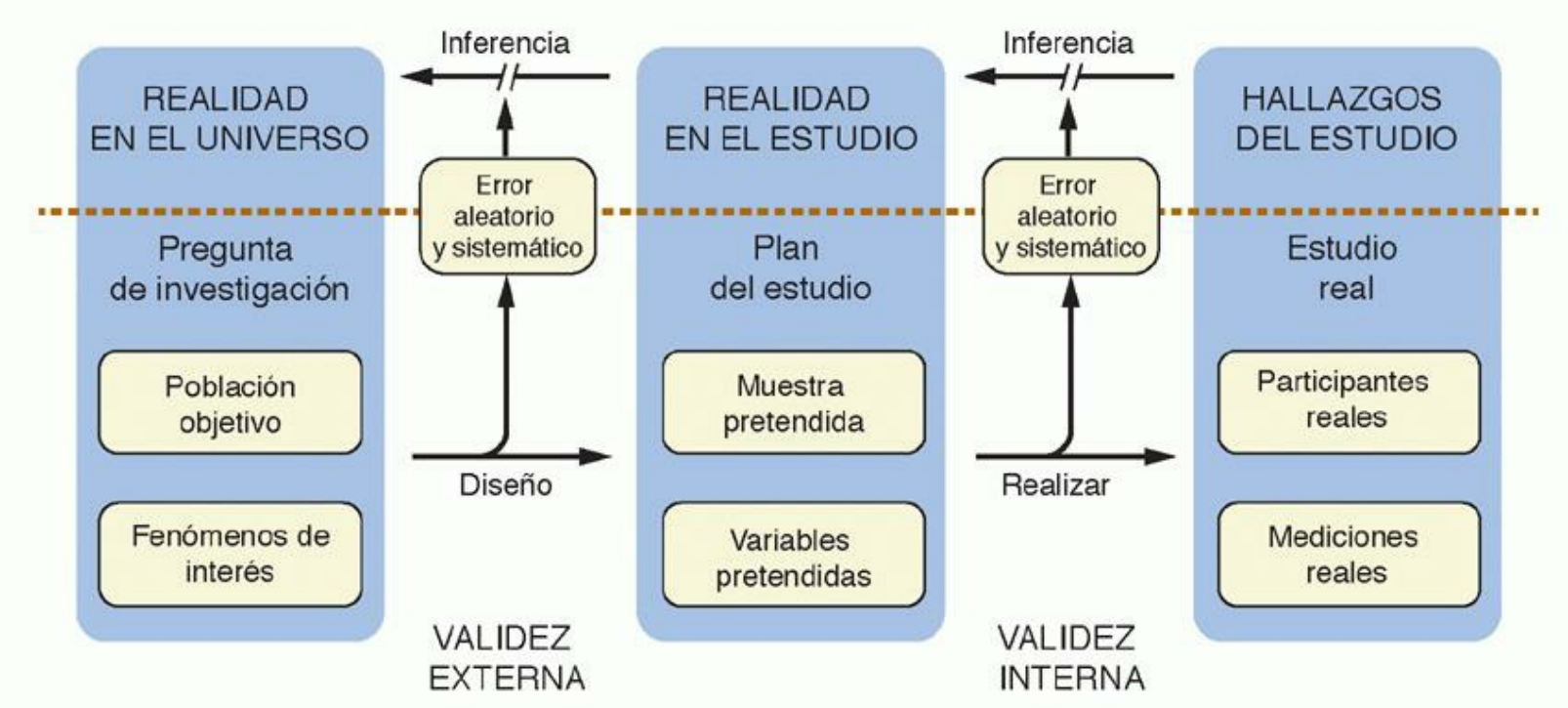
El error aleatorio es un resultado erróneo debido al azar: fuentes de variación con la misma probabilidad de alterar las mediciones del estudio en una u otra dirección. Si la frecuencia real de consumo diario de complementos de aceite de pescado en los varios centenares de pacientes de 50 a 69 años de edad con CI en la consulta del investigador es del 20 %, una muestra bien diseñada de 100 pacientes de esa población podría contener exactamente 20 pacientes que consumen esos complementos. Sin embargo, es más probable que la muestra contenga un número próximo, como 18, 19, 21 o 22. En ocasiones, el azar produciría una cifra notablemente diferente, como 12 o 28. Entre las diversas técnicas para reducir la influencia del error aleatorio (cap. 4), la más sencilla consiste en aumentar el tamaño de la muestra. El uso de una muestra de mayor tamaño disminuye la probabilidad de que se produzca un resultado muy erróneo al aumentar la precisión del cálculo: el grado en que la prevalencia observada se aproxima al 20 % cada vez que se obtiene una muestra.

El error sistemático es un resultado erróneo debido al sesgo: fuentes de variación que alteran los hallazgos del estudio en una dirección. Un ejemplo es la decisión en la figura 1-2 de estudiar pacientes en la consulta del investigador, donde los patrones de tratamiento locales han respondido a su interés en el tema y sus compañeros tienen más probabilidad que otros médicos de recomendar el aceite de pescado. El aumento del tamaño de la muestra carece de efecto sobre el error sistemático. El mejor modo de aumentar la exactitud del cálculo (grado en que se aproxima al valor verdadero) es diseñar el estudio de tal modo que reduzca la magnitud de los diversos sesgos. De manera alternativa, el investigador puede buscar información adicional para evaluar la importancia de los posibles sesgos. Un ejemplo sería comparar los resultados con los de una segunda muestra de pacientes con CI obtenida de otro entorno, por ejemplo, examinar si los hallazgos de pacientes de este tipo atendidos en una consulta de cardiología son diferentes de los encontrados en pacientes de una consulta de atención primaria.

Los ejemplos de error aleatorio y error sistemático de los dos párrafos anteriores son componentes del error de muestreo, que amenaza las inferencias desde los participantes del estudio a la población. Ambos tipos de errores (aleatorio y sistemático) pueden contribuir también al error de medición, que amenaza las inferencias desde las mediciones del estudio a los fenómenos de interés. Un ejemplo del error de medición aleatorio es la variación de la respuesta cuando se administra al paciente el cuestionario de la dieta en varias ocasiones. Un ejemplo de error de medición sistemático es la infravaloración de la frecuencia del consumo de aceite de pescado debido a la falta de claridad en la forma de expresar la pregunta. En los capítulos 3 y 4 se presentan más estrategias para controlar todas estas fuentes de error.



▪FIGURA 1-4. Errores de investigación. Estos detalles, desarrollados de los recuadros de errores de las figuras 1-2 y 1-3 , muestran estrategias para controlar el error aleatorio y sistemático en las fases de diseño y realización del estudio.
Herramientas de imágenes



▪FIGURA 1-5. Fisiología de la investigación: cómo funciona.
Herramientas de imágenes

En la figura 1-5 se resumen los conceptos que se han presentado en las páginas anteriores. Lograr la respuesta correcta a la pregunta de investigación es cuestión de diseñar y ejecutar el estudio de un modo que reduzca al mínimo la magnitud de los errores de inferencia.

Volver al principio

▪ DISEÑO DEL ESTUDIO

Protocolo del estudio

El proceso de elaborar el plan del estudio comienza con la pregunta de la investigación, en una frase que especifique el principal factor predictivo y las variables de respuesta y la población. Se elaboran secuencialmente tres versiones del plan del estudio, cada una mayor y más detallada que la que le precede:

- Resumen del estudio (v. tabla 1-1 y apéndice 1-1). Este resumen del diseño en una página actúa como lista de comprobación estandarizada que recuerda al investigador que debe abordar todos los componentes. También es importante que la secuencia tenga una lógica ordenada que ayude a aclarar las ideas del investigador sobre el tema.

- Protocolo del estudio. Esta ampliación del resumen del estudio habitualmente varía de 5 a 15 páginas y se utiliza para planificar el estudio y solicitar la aprobación del CEI y financiación. A lo largo del libro se exponen las partes del protocolo, y se resumen en el capítulo 19.

- Manual operativo. Esta recopilación de instrucciones específicas sobre el procedimiento, los cuestionarios y otros materiales está diseñada para asegurar un método uniforme y normalizado para realizar el estudio con un buen control de calidad (caps. 4 y 17).

La pregunta de la investigación y el resumen del estudio deben redactarse en una fase inicial. Reflejar las ideas en un papel transforma las ideas vagas en planes específicos, y proporciona una base concreta para obtener opiniones de compañeros y asesores. Hacerlo constituye un reto (es más fácil hablar de las ideas que escribirlas), pero las recompensas serán un inicio más rápido y un proyecto mejor.

En el apéndice 1 se presenta un ejemplo de un resumen de un estudio. Este resumen de una página se refiere más a la anatomía de la investigación (v. tabla 1-1) que a su fisiología (v. fig. 1-5), por lo que el investigador debe recordar los errores que pueden producirse en el momento de obtener inferencias, desde las medidas de la muestra en estudio hasta los fenómenos de interés en la población. Las virtudes y los problemas de un estudio pueden revelarse al considerar explícitamente en qué difiere la pregunta a la que es probable que responda el estudio de la pregunta de la investigación, según los planes para obtener participantes y realizar mediciones, y según los posibles problemas de ejecución.

Con el resumen del estudio en la mano y las inferencias propuestas en la mente, el investigador puede pasar a los detalles de su protocolo. Aquí se incluye la obtención de opiniones de compañeros, el bosquejo de métodos de inclusión de pacientes y de medición específicos, la consideración de la idoneidad científica y ética, la modificación de la pregunta del estudio y de su resumen cuando sea necesario, las pruebas preliminares para la inclusión de pacientes y los métodos específicos de medición, la realización de más cambios, la obtención de más opiniones, etc. Este proceso iterativo constituye la naturaleza del diseño de la investigación y será el tema del resto de este libro.

Concesiones

Lamentablemente, los errores son una parte inherente de todos los estudios. El principal tema es si estos errores serán lo suficientemente grandes como para cambiar las conclusiones de un modo importante. Cuando diseña un estudio, el investigador se encuentra en una posición muy parecida a la de un miembro de un sindicato negociando un nuevo contrato. Este empieza con una lista de deseos: menos horas, más dinero, mejora de la asistencia sanitaria, etc. Luego debe hacer concesiones, conservando las cosas que son más importantes y renunciando a las que no son esenciales o realistas. Al final de las negociaciones hay un paso fundamental: mira el mejor contrato que puede negociar y decide si se ha vuelto tan malo que ya no merece la pena.

El mismo tipo de concesiones debe hacer un investigador cuando transforma la pregunta de la investigación en el plan del estudio y considera los posibles problemas de su ejecución. A un lado se encuentran los problemas de validez interna y externa; al otro, la viabilidad. A veces, se omite el último paso esencial del negociador del sindicato. Una vez que se ha formulado el plan del estudio, el investigador debe decidir si aborda adecuadamente la pregunta de la investigación y si puede realizarse con unos niveles de error aceptables. A menudo la respuesta es no, y será necesario empezar el proceso de nuevo. Pero ¡anímese! Los buenos científicos se distinguen no tanto por sus ideas de investigación uniformemente buenas como por su entusiasmo en abandonar las que no funcionan y pasar a otras mejores.

Volver al principio

▪ RESUMEN

- La anatomía de la investigación es el conjunto de elementos tangibles que conforman el plan del estudio: la pregunta de la investigación y su importancia, y el diseño, los participantes en el estudio y los métodos de medición. El reto consiste en diseñar elementos que sean relativamente económicos y fáciles de realizar.
- La fisiología de la investigación es el modo en que se realiza el estudio. Los hallazgos de este se usan para obtener inferencias sobre lo que sucedió en la muestra de estudio (validez interna) y sobre acontecimientos fuera de este (validez externa). El reto aquí consiste en diseñar y ejecutar un plan de estudio con un control adecuado sobre dos importantes amenazas a estas inferencias: el error aleatorio (azar) y el error sistemático (sesgo).
- Al diseñar un estudio el investigador puede encontrar que es útil considerar la figura 1-5 , las relaciones entre la pregunta de la investigación (lo que se quiere responder), el plan del estudio (lo que el estudio pretende responder) y el estudio real (lo que el estudio realmente responderá, según los errores de ejecución que pueden preverse).
- Una buena forma de desarrollar el plan del estudio consiste en empezar con una versión de la pregunta de la investigación en una frase que especifique las variables principales y la población, y ampliarla en un resumen de una página que describa los elementos en una secuencia normalizada. Más adelante se ampliará en el protocolo y el manual operativo.
- Se necesita un buen juicio por parte del investigador y el consejo de los colaboradores para las numerosas concesiones que se deben hacer, y para determinar la viabilidad general del proyecto.

Volver al principio

[1]Las variables predictivas en ocasiones se denominan variables independientes, y las variables de respuesta variables dependientes, aunque el significado de estos términos es menos evidente y preferimos evitar su uso.

APÉNDICE 1Resumen de un estudio
NA

Este es el plan del estudio en una página de un proyecto realizado por Valerie Flaherman, MD, MPH, que comenzó cuando era residente

de pediatría general en UCSF. Para la mayoría de los investigadores principiantes es más fácil iniciar estudios de observación, aunque en este caso era posible un estudio clínico aleatorizado de un tamaño y un ámbito moderados, el único diseño que permitía abordar adecuadamente la pregunta de la investigación, y en último término tuvo éxito (v. en la publicación de Flaherman y cols.[1] los hallazgos que, si se confirman, podrían alterar las políticas de cómo iniciar mejor la lactancia materna).

▪ TÍTULO: EFECTO DEL USO PRECOZ DE UNA FÓRMULA LIMITADA SOBRE LA LACTANCIA MATERNA

Pregunta de la investigación

En recién nacidos a término que han perdido ≥ 5 % del peso al nacimiento antes de las 36 h de edad, ¿la administración de 10 cm de fórmula con una jeringuilla después de cada toma de lactancia materna antes del inicio de la producción de leche madura aumenta la probabilidad de éxito en la lactancia materna posterior?

Importancia

- El volumen de la leche materna es bajo hasta que comienza la producción de leche madura 2-5 días después del parto.
- Algunas madres se preocupan si el inicio de la producción de leche madura es tardío y su hijo pierde mucho peso, y todo ello las lleva a abandonar la lactancia materna en la primera semana. Una estrategia que aumentará la proporción de madres que tienen éxito en la lactancia materna produciría muchos beneficios sanitarios y psicosociales para la madre y el niño.
- En estudios de observación se ha encontrado que la alimentación con fórmula en los primeros días tras el parto se asocia a una disminución de la lactancia materna. Aunque esto se podría deber a la confusión por la indicación (v. cap. 9), este hallazgo ha llevado a directrices de la OMS y de los CDC dirigidas a reducir el uso de fórmula durante la hospitalización del parto.
- Sin embargo, la combinación de una pequeña cantidad de fórmula con la lactancia materna y el consejo médico podría hacer que la experiencia de lactancia materna temprana fuera más positiva y aumentaría la probabilidad de éxito. Hace falta un estudio clínico para evaluar los posibles beneficios y perjuicios de esta estrategia.

Diseño del estudio

Estudio controlado y aleatorizado sin enmascaramiento, con ocultación para la evaluación de los criterios de valoración.

Participantes

- Criterios de inclusión: recién nacidos a término sanos de 24-48 h de edad que hayan perdido ≥ 5 % del peso corporal en las primeras 36 h tras el nacimiento.
- Diseño del muestreo: muestra consecutiva de pacientes que dieron su consentimiento en dos centros médicos académicos del norte de California.

Variable predictiva, asignada aleatoriamente, pero sin enmascaramiento

- Testigo: se enseñan a los progenitores técnicas para calmar al lactante.
- Intervención: se enseña a los progenitores a administrar con una jeringuilla 10 cm de fórmula después de cada toma de lactancia materna hasta el inicio de la producción de leche madura.

Variables de respuesta evaluadas con enmascaramiento de la asignación

- Cualquier toma de fórmula al cabo de 1 semana y de 1, 2 y 3 meses.
- Cualquier toma de lactancia materna al cabo de 1 semana y de 1, 2 y 3 meses.
- Valor mínimo del peso.

Hipótesis nula principal

La administración temprana limitada de fórmula no afecta a la proporción de mujeres que alimentan a su hijo con lactancia natural a los 3 meses.

BIBLIOGRAFÍA

1. Flaherman VJ, Aby J, Burgos AE, et al. Effect of early limited formula on duration and exclusivity of breastfeeding in at-risk infants: an RCT. Pediatrics, en prensa.

CAPÍTULO 2 Elaboración de la pregunta de la investigación y desarrollo del plan del estudio

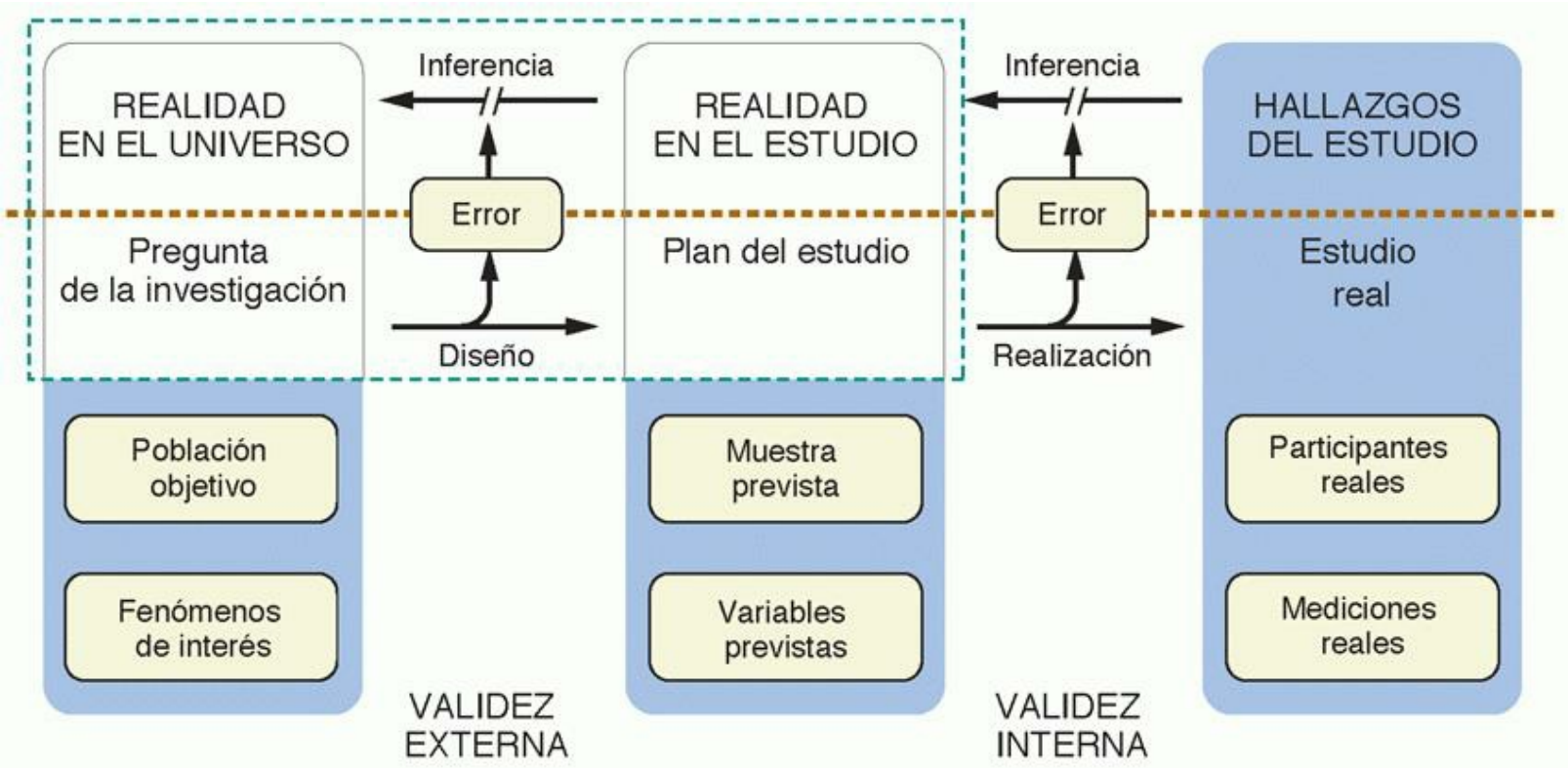
Steven R. Cummings
Warren S. Browner
Stephen B. Hulley

La pregunta de la investigación es la incertidumbre que el investigador quiere resolver realizando el estudio. No hay escasez de buenas preguntas de investigación, e incluso si logramos dar respuesta a algunas, seguimos rodeados por otras. Por ejemplo, en algunos ensayos clínicos se ha establecido que los tratamientos que bloquean la síntesis de estradiol (inhibidores de la aromatasa) reducen el riesgo de cáncer de mama en mujeres que han sufrido cáncer en estadio inicial (1). Pero esto ha llevado a nuevas preguntas: ¿durante cuánto tiempo debe continuarse el tratamiento?; ¿previene este tratamiento el cáncer de mama en pacientes con mutaciones de *BRCA1* y *BCRA2*?, y ¿cuál es el mejor modo de prevenir la osteoporosis, que es un efecto adverso de estos fármacos? Aparte de eso, se plantean dudas sobre la prevención primaria: ¿son estos fármacos eficaces y seguros para la prevención del cáncer de mama en mujeres sanas?

El reto en la búsqueda de una pregunta de la investigación es la definición de una pregunta importante que se pueda transformar en un plan de estudio viable y válido. En este capítulo se presentan estrategias para lograrlo (fig. 2-1).

▪ ORÍGENES DE UNA PREGUNTA DE INVESTIGACIÓN

Para un investigador con experiencia, las mejores preguntas de investigación suelen surgir de los hallazgos y problemas que ha observado en sus propios estudios anteriores, así como en los de otros investigadores que trabajan en el mismo campo. Un nuevo investigador no ha desarrollado todavía esta base de experiencia. Aunque una perspectiva nueva es útil a veces, porque permite que una persona creativa plantee nuevos enfoques a los problemas antiguos, la falta de experiencia es, sobre todo, un impedimento.



▪ FIGURA 2-1. Este capítulo se centra en el área dentro de la línea verde discontinua, el reto de elegir una pregunta de la investigación que tenga interés y que se pueda abordar con un plan de estudio viable.
Herramientas de imágenes

Una buena forma de comenzar es clarificar la diferencia entre una pregunta de la investigación y un interés de la investigación. Tome en consideración esta pregunta de la investigación:

- La participación en sesiones de asesoramiento de grupo, ¿reduce la probabilidad de violencia doméstica en mujeres que han emigrado recientemente desde Centroamérica?

Esta pregunta la podría plantear alguien cuyo interés de investigación se refiere a la eficacia del asesoramiento de grupo, o a la prevención de la violencia doméstica, o a la mejora de la salud en inmigrantes recientes. La distinción entre preguntas de la investigación e intereses de la investigación es importante, porque puede ocurrir que la pregunta de la investigación específica no se pueda transformar en un plan de estudio viable, aunque el investigador puede seguir abordando su interés de la investigación haciendo una pregunta

diferente.

Por supuesto, es imposible formular una pregunta de la investigación si no se está seguro siquiera del interés de la investigación (aparte de saber qué se supone que quiere uno). Si usted se encuentra en esta situación, no está solo: muchos nuevos investigadores todavía no han descubierto un tema que les interese y que sea susceptible de un plan de estudio que puedan diseñar. Puede empezarse considerando qué tipos de estudios de investigación han estimulado su interés cuando los ve en una revista médica. O tal vez se haya sentido preocupado por un paciente específico cuyo tratamiento parecía inadecuado o incorrecto: ¿qué se podría haber realizado de manera diferente para mejorar el resultado? O tal vez uno de los médicos ayudantes le haya comentado que la hipopotasemia siempre producía una sed profunda, y otro dijo lo contrario, de una manera igual de dogmática.

Conocer a fondo la bibliografía

Es muy importante conocer a fondo la bibliografía publicada en un campo de estudio: la erudición es un precursor necesario de una buena investigación. Un investigador nuevo debe realizar una exhaustiva búsqueda de la bibliografía publicada en las áreas relativas a la pregunta de la investigación y debe leer con sentido crítico los principales artículos originales. Llevar a cabo una revisión sistemática constituye un importante paso siguiente para adquirir experiencia en un campo de investigación, y la revisión de la bibliografía subyacente puede servir de introducción a las propuestas para obtener subvenciones y al artículo de investigación. Puede que los avances recientes sean conocidos por los investigadores activos en un campo concreto mucho antes de ser publicados. Así pues, conocer a fondo un tema conlleva participar en reuniones y establecer relaciones con expertos en ese tema.

Estar alerta ante nuevas ideas y técnicas

Además de la bibliografía médica como fuente de ideas para temas de investigación, es útil asistir a conferencias en las que se presentan trabajos recientes. Al menos tan importantes como las presentaciones son las oportunidades de entablar conversaciones informales con otros científicos en las sesiones de pósteres y durante los descansos. Un nuevo investigador que supere su timidez y se dirija a un conferenciante durante el descanso para tomar café puede encontrar que es una experiencia muy provechosa, y en ocasiones tendrá un nuevo colega con más experiencia. Incluso, aún más, en caso de conocer de antemano la participación de un conferenciante que va a ser especialmente importante, puede que merezca la pena leer sus publicaciones recientes y contactar con él por adelantado para concertar una reunión durante la conferencia.

Una actitud escéptica ante creencias comunes puede ser un estímulo para formular buenas preguntas de investigación. Por ejemplo, en general se pensaba que las laceraciones que van más allá de la dermis precisaban suturas que aseguraran una rápida curación y unos resultados estéticos satisfactorios. Sin embargo, Quinn y cols. describieron su experiencia personal, y en series de casos se vio que las heridas de tamaño moderado se curan por sí mismas independientemente de si se han aproximado o no sus bordes (2). Llevaron a cabo un ensayo clínico aleatorizado en el que los pacientes con laceraciones de menos de 2 cm de longitud en las manos fueron tratados con irrigación con agua corriente y un apósito con antibióticos durante 48 h. A un grupo se le asignó aleatoriamente a sutura de las heridas, y en el otro grupo no se suturaron. El grupo de las suturas recibió un tratamiento más doloroso y más prolongado en el servicio de urgencias, aunque la evaluación con enmascaramiento mostró que los tiempos de curación y los resultados estéticos eran similares. Esto se ha convertido actualmente en el abordaje estándar que se utiliza en la práctica clínica.

La aplicación de nuevas tecnologías a menudo genera nuevas ideas y preguntas sobre problemas clínicos habituales, que, a su vez, pueden generar nuevos paradigmas (3). Los avances en el diagnóstico por la imagen y en técnicas de análisis genéticos y moleculares, por ejemplo, han generado estudios de investigación aplicada que han llevado a nuevos tratamientos y pruebas que han modificado la medicina clínica. Del mismo modo, la obtención de un nuevo concepto, tecnología o hallazgo en un campo y su aplicación a un problema de un campo diferente pueden conducir a la formulación de buenas preguntas de investigación. Por ejemplo, una densidad ósea baja es un factor de riesgo de fracturas. Los investigadores aplicaron esta tecnología a otros problemas y observaron que las mujeres con densidad ósea baja tienen mayor velocidad de deterioro cognitivo (4), lo cual estimuló la investigación de factores, como concentraciones endógenas bajas de estrógenos, que podrían llevar a la pérdida de hueso y de memoria.

Mantener viva la imaginación

La observación meticulosa de los pacientes ha dado lugar a muchos estudios descriptivos y es una fuente abundante de preguntas para la investigación. La docencia es también una excelente fuente de inspiración; a menudo surgen ideas para nuevos estudios mientras se preparan presentaciones o durante debates con estudiantes curiosos. Debido a que no suele disponerse de tiempo suficiente para desarrollar esas ideas en el acto, es útil conservarlas en un archivo informatizado o en una agenda para recurrir a ellas en el futuro.

La creatividad desempeña un importante papel en el proceso de formular nuevas preguntas de investigación, imaginar nuevos métodos para abordar cuestiones antiguas y jugar con las ideas. Algunas ideas creativas vienen a la mente durante conversaciones informales con los colegas durante el almuerzo; otras surgen en la discusión de la investigación reciente o de las propias ideas en grupos pequeños.

Muchas inspiraciones son acontecimientos solitarios que brotan mientras se prepara una lectura, en la ducha, escudriñando en Internet, o simplemente sentándose y pensando. El miedo a la crítica o a parecer raro puede anular prematuramente las nuevas ideas. El truco está en colocar a la vista, claramente, un problema sin resolver y permitir que la mente vuele libremente alrededor de él. También es necesaria la tenacidad, volver a un problema molesto una y otra vez hasta que se alcance una solución.

Buscar un mentor y trabajar con él

No hay nada que sustituya a la experiencia a la hora de orientar las muchas opiniones que intervienen en la formulación de una pregunta para la investigación y la elaboración de un plan de estudio. Por lo tanto, una estrategia esencial para un nuevo investigador es convertirse en aprendiz de un mentor con experiencia que disponga de tiempo e interés para trabajar con él con regularidad.

Un buen mentor estará disponible para tener reuniones frecuentes y conversaciones informales, animará las ideas creativas, proporcionará conocimientos procedentes de su experiencia, se asegurará de que tiene tiempo para dedicarlo a la investigación, abrirá puertas a oportunidades de financiación y contactos, animará al desarrollo de trabajos independientes, y colocará el nombre del nuevo investigador en primer lugar en las subvenciones y publicaciones siempre que sea procedente. A veces es deseable contar con más de un mentor, y que todos ellos representen disciplinas diferentes. Unas buenas relaciones de este tipo también pueden proporcionar los recursos tangibles que se necesitan: espacio en consulta, acceso a poblaciones clínicas, conjuntos de datos y bancos de muestras, laboratorios especializados, recursos económicos y un equipo de investigación.

Por otro lado, un mal mentor puede ser una barrera. Un mentor puede perjudicar la carrera de un nuevo investigador, por ejemplo, atribuyéndose el mérito de los hallazgos que se originen en el trabajo del nuevo investigador o asumiendo el papel principal en su publicación o presentación. Con más frecuencia, muchos mentores simplemente están demasiado ocupados o distraídos para prestar atención a las necesidades del nuevo investigador. En cualquier caso, una vez que las discusiones con el mentor hayan resultado estériles, recomendamos encontrar una forma de pasar a un asesor más adecuado, tal vez implicando en las negociaciones a un compañero con más experiencia que sea neutral. El cambio de mentor puede ser difícil, lo cual pone de manifiesto la importancia de elegir un buen mentor desde el principio; tal vez sea la *decisión única más importante* que toma un nuevo investigador.

Su mentor le puede ofrecer una base de datos y pedirle que venga con una pregunta de la investigación. En esa situación es importante identificar: *1)* la superposición entre lo que hay en la base de datos y sus propios intereses de investigación, y *2)* la calidad de la base de datos. Si no hay suficiente superposición o si los datos contienen errores irrevocables, debe encontrarse una forma de pasar a otro proyecto.

Volver al principio

▪ CARACTERÍSTICAS DE UNA BUENA PREGUNTA DE LA INVESTIGACIÓN

Las características de una pregunta de investigación que lleve a un buen plan de estudio es que sea factible, interesante, novedosa, ética y relevante (palabras que forman el acrónimo FINER; tabla 2-1).

Factible

Es mejor conocer los límites y problemas prácticos del estudio de una pregunta en fases tempranas de la investigación, antes de invertir mucho tiempo y esfuerzo en líneas impracticables.

- Número de participantes. Muchos estudios no logran los objetivos propuestos por no poder contar con suficientes participantes. Puede ser muy útil hacer pronto un cálculo preliminar de los requisitos del tamaño de la muestra del estudio (cap. 6), junto con un cálculo del número de participantes de los que probablemente se dispondrá para el estudio, el número de los que serían excluidos o rechazarían participar, y el número de los que se perderían durante el seguimiento. Incluso una planificación meticulosa produce muchas veces cálculos que son excesivamente optimistas, y el investigador debe asegurarse de que haya suficientes participantes idóneos y dispuestos. A veces es necesario realizar un estudio piloto o una revisión de las historias clínicas para estar seguro. Si parece que el número de participantes es insuficiente, el investigador puede considerar varias estrategias: ampliación de los criterios de inclusión, eliminación de criterios de exclusión innecesarios, alargamiento del plazo de tiempo para incluir participantes, adquisición de otras fuentes de participantes, desarrollo de métodos de medición más precisos, invitación a compañeros para que participen en un estudio multicéntrico y utilización de un diseño del estudio diferente.

- Experiencia técnica. Los investigadores deben tener la capacidad, el equipo y la experiencia necesarios para diseñar el estudio, incluir a los participantes, medir las variables, y gestionar y analizar los datos. Los consultores pueden ayudar a apuntalar aspectos técnicos con los que no estén familiarizados los investigadores, pero, para áreas importantes del estudio, es mejor contar con un

compañero con experiencia que participe intensamente como coinvestigador; por ejemplo, es sensato contar con un estadístico como miembro del equipo de investigación desde el principio del proceso de planificación. Es mejor utilizar métodos familiares y ya establecidos, porque el proceso de elaborar nuevos métodos y aptitudes lleva tiempo y no es seguro. Cuando se precisa un nuevo método, como la medición de un nuevo biomarcador, debe buscarse la experiencia en la forma de llevar a cabo la innovación.

Factible

Número adecuado de participantes

Experiencia técnica adecuada

Asequible en cuanto a tiempo y recursos económicos

Manejable en cuanto al ámbito

Financiable

Interesante

La consecución de la respuesta intriga al investigador y a sus colegas

Novedosa

Proporciona nuevos hallazgos

Confirma, refuta o amplía hallazgos previos

Puede llevar a innovaciones en conceptos de salud y enfermedad, práctica médica o metodologías de investigación

Ética

Un estudio que aprobará el comité de ética de la investigación médica

Relevante

Con probabilidad de tener consecuencias significativas sobre el conocimiento científico, la práctica clínica o las políticas sanitarias

Puede influir en las perspectivas de la investigación futura

- Coste de tiempo y económico. Es importante calcular los costes de cada uno de los componentes del proyecto, teniendo en cuenta

que el tiempo y el dinero necesarios superarán generalmente las cantidades previstas al principio. Si los costes proyectados superan los fondos de los que se dispone, las únicas opciones serán considerar un estudio más barato o buscar nuevas fuentes de financiación. Reconocer pronto que un estudio es demasiado caro o que va a requerir demasiado tiempo puede hacer que se modifique o se abandone el plan antes de que se haya realizado un gran esfuerzo.

- Ámbito. A menudo surgen problemas cuando un investigador intenta conseguir demasiado, realizando muchas determinaciones en contactos repetidos con un gran grupo de participantes en un intento de responder a demasiadas preguntas. La solución estará en acortar el campo de estudio y centrarse solo en los objetivos más importantes. Muchos científicos piensan que es difícil abandonar la oportunidad de responder a preguntas accesorias interesantes, pero la recompensa puede ser lograr una mejor respuesta a la pregunta principal.

- Posibilidad de financiación. Pocos investigadores disponen de recursos personales o institucionales para financiar sus propios proyectos de investigación, particularmente si se debe incluir y seguir a pacientes, o si se deben realizar mediciones costosas. La propuesta de investigación con el diseño más elegante no será viable si no hay quien la pague. En el capítulo 19 se analiza la búsqueda de fuentes de financiación.

Interesante

Un investigador puede tener muchos motivos para perseguir una pregunta de la investigación concreta: porque le proporcionará dinero, porque es un siguiente paso, lógico o importante, en su carrera profesional, o porque resulta interesante buscar la verdad del tema. Nos gusta esta última razón; crece a medida que se ejercita, y proporciona la intensidad de esfuerzo necesaria para superar los numerosos obstáculos y frustraciones del proceso de investigación. Sin embargo, es prudente confirmar que no es usted el único que encuentra que una pregunta es interesante. Hable con mentores, expertos externos y representantes de posibles agencias de financiación, como los encargados de proyectos del NIH, antes de emplear demasiada energía en desarrollar un plan de investigación o una propuesta de financiación que los compañeros y entidades financieras pueden considerar carente de interés.

Novedosa

Una buena investigación clínica aportará información novedosa. Un estudio que simplemente reitere lo que ya está establecido no merece el esfuerzo ni el coste, y es improbable que reciba financiación. La novedad de un estudio propuesto puede determinarse mediante una revisión exhaustiva de la bibliografía, la consulta con expertos familiarizados en la investigación en curso no publicada y la búsqueda de resúmenes de proyectos en el campo de interés que han sido financiados consultando la página web NIH Research Portfolio Online Reporting Tools (RePORT) (http://report.nih.gov/categorical_spending.aspx.). Las revisiones de estudios enviados al NIH dan un peso considerable a si un estudio propuesto es innovador (5), de modo que un resultado positivo podría cambiar paradigmas de investigación o la práctica clínica mediante el uso de nuevos conceptos, métodos o intervenciones (cap. 19). Aunque la novedad es un criterio importante, no es necesario que una pregunta de investigación sea totalmente original: puede merecer la pena preguntarse si puede repetirse una observación anterior, si los hallazgos en una población se aplican también a otras, o si un nuevo método de medición puede aclarar la relación entre factores de riesgo conocidos y una enfermedad. Un estudio de confirmación será particularmente útil si evita los puntos débiles de estudios anteriores o si el resultado a confirmar era inesperado.

Ética

Una buena pregunta de investigación debe ser ética. Si el estudio presenta riesgos físicos inaceptables o invasión de la privacidad (cap. 14), el investigador debe buscar otras vías para responder a la pregunta. Si existen dudas sobre si el estudio es ético, será útil comentarlo al principio con un representante del comité de ética de la investigación (CEI).

Relevante

Un buen método para decidir sobre la importancia es imaginar los diversos resultados que pueden suceder y considerar cómo cada posibilidad podría hacer avanzar el conocimiento científico, influir en directrices prácticas y políticas sanitarias, u orientar a investigaciones posteriores. Los revisores del NIH insisten en la relevancia de un estudio propuesto: la importancia del problema, en qué medida el proyecto mejorará el conocimiento científico, y de qué manera el resultado modificará conceptos, métodos o servicios clínicos.

Volver al principio

▪ DESARROLLO DE LA PREGUNTA DE LA INVESTIGACIÓN Y DEL PLAN DE ESTUDIO

Resulta muy útil escribir la pregunta de investigación y un resumen breve (de una página) del plan del estudio en una etapa inicial (apéndice 1). Para ello se necesita algo de autodisciplina, pero obliga al investigador a aclarar sus ideas sobre el plan y a descubrir

problemas específicos que requieren atención. El resumen proporciona también una base para las sugerencias específicas de los compañeros.

Problemas y abordajes

Dos abordajes complementarios a los problemas que supone la formulación de una pregunta de investigación merecen especial atención.

La primera es la importancia de obtener un buen asesoramiento. Recomendamos un equipo de investigación que incluya representantes de cada una de las principales disciplinas del estudio, y que incluya al menos un científico con experiencia. Además, es una buena idea consultar con especialistas que puedan orientar el descubrimiento de investigaciones anteriores sobre el tema, así como la elección y el diseño de las técnicas de medición. A veces lo hará un experto local, pero a menudo resulta útil contactar con personas de otros centros que hayan publicado algún trabajo acerca del tema. Un nuevo investigador puede verse intimidado por la idea de escribir o llamar a alguien a quien solo conoce como autor en la revista *Journal of the American Medical Association*, pero la mayoría de los científicos responden favorablemente a las peticiones de estos consejos.

El segundo abordaje es permitir que el plan de estudio surja gradualmente de un proceso iterativo de realización de cambios incrementales en el diseño del estudio, estimación del tamaño de la muestra, revisión con los compañeros, estudio previo de las características principales y revisión. Una vez que se ha especificado el resumen del estudio en una página, la revisión formal por los compañeros habitualmente llevará a mejoras importantes. A medida que el protocolo va tomando forma, estudios piloto de la disponibilidad y la voluntad de números suficientes de participantes pueden llevar a cambios del plan de inclusión. La prueba de imagen preferida puede llegar a ser prohibitivamente cara, por lo que hay que buscar una alternativa más barata.

Preguntas principales y secundarias

Muchos estudios tienen más de una pregunta de investigación. Los experimentos a veces abordan el efecto de la intervención hacia algo más de un resultado; por ejemplo, se diseñó la *Women's Health Initiative* para determinar si la reducción del consumo de grasa en la dieta reduciría el riesgo de cáncer de mama, pero una hipótesis secundaria importante fue examinar el efecto sobre los episodios coronarios (5). Casi todos los estudios de cohortes y de casos y testigos analizan varios factores de riesgo para cada resultado. La ventaja de diseñar un estudio con varias preguntas de investigación es la eficacia que puede lograrse, porque pueden surgir varias respuestas a partir de un solo estudio. Los inconvenientes son la mayor complejidad del diseño y de la ejecución del estudio, y la obtención de inferencias estadísticas cuando existen múltiples hipótesis (cap. 5). Una estrategia sensata es establecer una única pregunta de investigación principal alrededor de la cual se centra el plan de estudio y el cálculo del tamaño de la muestra, añadiéndose preguntas de la investigación secundarias sobre otras variables predictivas o de respuesta que pueden generar también conclusiones valiosas.

Volver al principio

▪ INVESTIGACIÓN APLICADA

La investigación aplicada se refiere a los estudios que buscan la manera de llevar los hallazgos desde la «torre de marfil» al «mundo real», de garantizar que la creatividad científica tenga consecuencias favorables sobre la salud pública. La investigación aplicada (6) se presenta en dos modalidades principales (fig. 2-2):

- Aplicación de hallazgos científicos básicos de investigación de laboratorio a estudios clínicos de pacientes (a veces abreviado como investigación T1), y
- Aplicación de los hallazgos de estos estudios clínicos para alterar prácticas sanitarias en la comunidad (a veces abreviado como investigación T2).

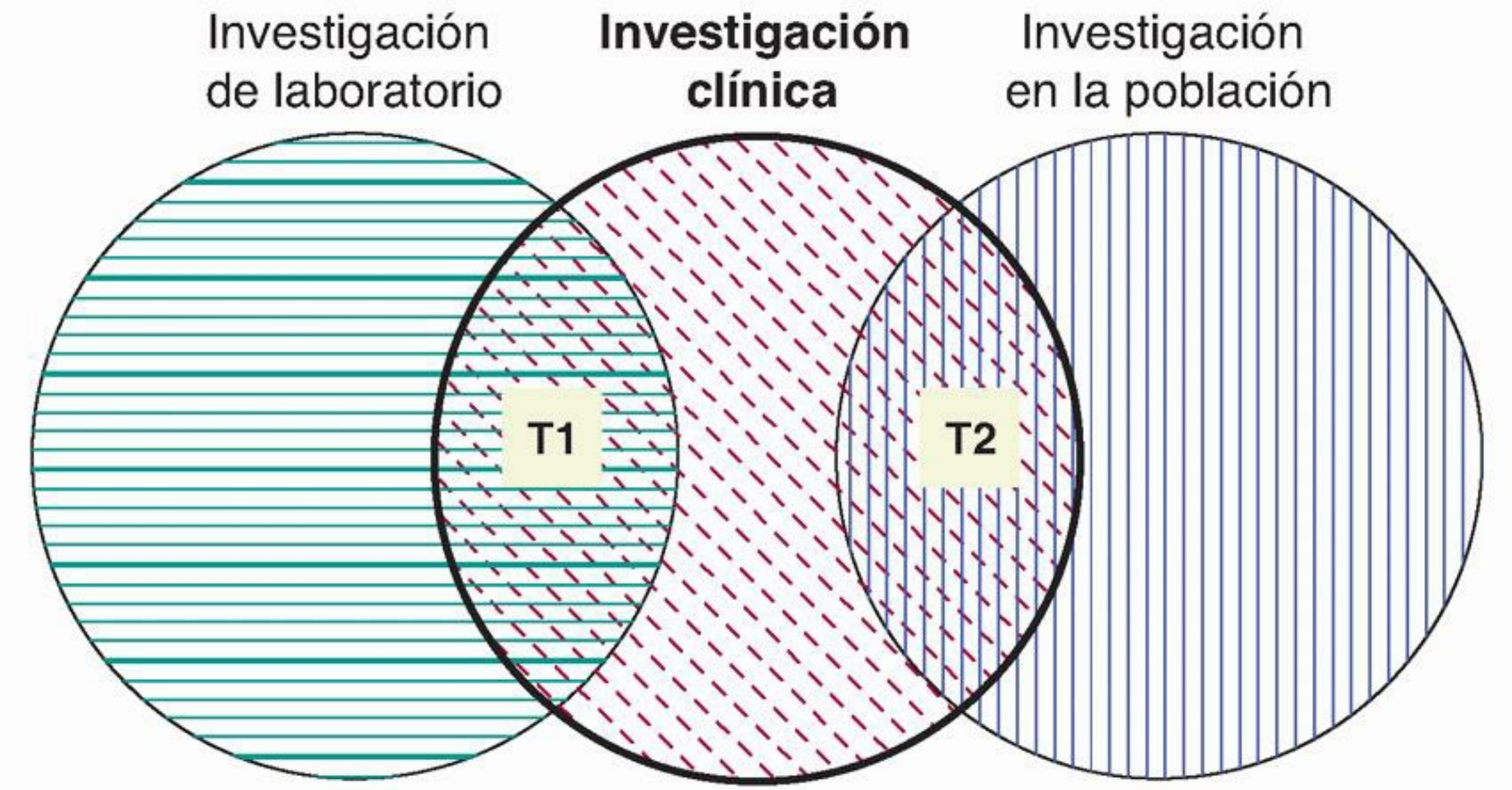
Ambas formas de investigación aplicada requieren la identificación de una oportunidad de «aplicación». Tal como un traductor literario necesita, en primer lugar, encontrar una novela o un poema que merezca ser traducido, un investigador aplicado debe encontrar primero un hallazgo científico o una nueva tecnología que pueda tener consecuencias importantes sobre la investigación clínica, la práctica clínica o la salud pública. Entre las estrategias para hacer esta difícil elección, puede ser útil prestar atención a los compañeros cuando hablan sobre sus hallazgos más recientes, a las presentaciones en congresos nacionales sobre nuevos métodos, y a la especulación sobre mecanismos en artículos publicados.

Aplicación desde el laboratorio a la práctica clínica (T1)

Han surgido múltiples herramientas para las investigaciones clínicas, entre ellas secuenciado del ADN, matrices de expresión génica,

técnicas de imagen molecular y proteómica. Desde el punto de vista de un investigador clínico, estas nuevas mediciones, tecnologías o resultados de pruebas no tienen ningún aspecto que sea diferente desde el punto de vista epidemiológico. Los capítulos sobre mediciones serán útiles para planificar estudios en los que intervengan estos tipos de determinaciones (cap. 4), así como también lo será el consejo sobre el diseño del estudio (caps. 7-12), las muestras de población (cap. 3) y el tamaño de la muestra (cap. 6). El problema de la comprobación de múltiples hipótesis (cap. 5) será especialmente importante para la genómica y otras «-ómicas».

En comparación con la investigación clínica ordinaria, para ser un investigador aplicado T1 con éxito a menudo se debe contar con una serie de capacidades adicionales o identificar a un colaborador que las tenga. La investigación del laboratorio a la práctica clínica precisa un conocimiento exhaustivo de los conceptos básicos subyacentes. Aunque muchos investigadores clínicos piensan que pueden dominar este conocimiento (como muchos investigadores de laboratorio creen que la investigación clínica no precisa ninguna formación especial), en realidad las capacidades apenas se superponen. Por ejemplo, suponga que un científico básico ha identificado un gen que afecta al ritmo circadiano en los ratones. Un investigador clínico con experiencia en el estudio del sueño tiene acceso a un estudio de cohortes con datos sobre los ciclos del sueño y un banco de ADN almacenado, y desea estudiar si existe una asociación entre las variantes del homólogo humano de ese gen y el sueño. Con el fin de proponer un estudio T1 que contemple esa asociación, necesita colaboradores que estén familiarizados con ese gen y con las ventajas y limitaciones de los diversos métodos de genotipificación.



▪ FIGURA 2-2. La investigación aplicada es el componente de la investigación clínica que interactúa con la investigación científica básica (área sombreada T1) o con la investigación en la población (área sombreada T2).

Herramientas de imágenes

Del mismo modo, imagine que un investigador de laboratorio ha descubierto un patrón de expresión génica característico en muestras de biopsias tisulares de pacientes con cáncer de mama. No debe proponer un estudio de su uso como prueba para predecir el riesgo de recurrencia del cáncer de mama sin colaborar con alguien que conozca la importancia de los aspectos de la investigación clínica, como la fiabilidad de las pruebas, el muestreo y el enmascaramiento, así como los efectos de la probabilidad previa de enfermedad sobre la posibilidad de aplicación de su descubrimiento. Una buena investigación aplicada necesita experiencia en más de un campo. Así pues, un equipo de investigación que esté interesado en estudiar un nuevo fármaco necesita científicos familiarizados con la biología molecular, la farmacocinética, la farmacodinámica, los ensayos clínicos de fase I y II, y la práctica actual en el correspondiente campo de la medicina.

Investigación desde la práctica clínica a las poblaciones (T2)

Los estudios que pretenden aplicar hallazgos de estudios cínicos a poblaciones más diversas y de mayor tamaño precisan con frecuencia experiencia en la identificación de grupos de alto riesgo o a los que se ha prestado poca atención, en la comprensión de la diferencia entre detección selectiva y diagnóstico, y en el conocimiento de cómo llevar a cabo cambios en los sistemas de prestación de asistencia sanitaria. A un nivel práctico, este tipo de investigación suele necesitar el acceso a grandes grupos de pacientes (o médicos), como los que intervienen en planes sanitarios o consultas grandes. El apoyo y el consejo del jefe del departamento, del jefe del equipo médico de

un hospital afiliado, del director de una organización de asistencia gestionada o de un representante de una organización comunitaria pueden ser útiles al planificar estos estudios.

Algunos investigadores toman un atajo al realizar este tipo de investigación aplicada, ampliando un estudio de su propia consulta estudiando pacientes de las consultas de sus compañeros (es decir, una clínica en un centro médico académico) en lugar de contar con profesionales de la comunidad. Esto es un poco como traducir a Aristófanes al griego moderno: seguirá sin ser muy útil para lectores que hablan castellano. El capítulo 18 hace hincapié en la importancia de implicarse lo más posible con la comunidad.

La verificación de los hallazgos de la investigación en poblaciones mayores muchas veces precisa la adaptación de los métodos a las organizaciones. Por ejemplo, en un estudio sobre si una nueva dieta aplicada en la consulta y un programa de ejercicio serán eficaces en la comunidad, tal vez no sea posible asignar a los pacientes concretos de forma aleatoria. Una solución sería asignar aleatoriamente las consultas de los médicos. Para ello se necesitará la colaboración de un experto en muestreo por conglomerados y análisis por conglomerados. Muchos proyectos de investigación T2 dirigidos a mejorar la asistencia médica utilizan variables de «proceso» indirectas como resultados. Por ejemplo, si los ensayos clínicos han establecido que un nuevo tratamiento reduce la mortalidad por sepsis, un estudio de investigación aplicada que compare dos programas para la aplicación y la promoción del uso del nuevo tratamiento no tendría por qué tener la mortalidad como criterio de valoración. Antes bien, simplemente se podría comparar el porcentaje de pacientes con sepsis que han recibido el nuevo tratamiento. El desplazamiento de la investigación desde contextos diseñados para la investigación a organizaciones diseñadas para la asistencia médica u otros fines precisa flexibilidad y creatividad en la aplicación de los principios que garantizan el rigor y la validez máximos posibles de los resultados del estudio.

Volver al principio

▪ RESUMEN

- Todos los estudios deben empezar con una pregunta de la investigación que aborde lo que el investigador desearía conocer. El objetivo es encontrar una que pueda desarrollarse en un buen plan de estudio.
- La erudición es esencial para formular preguntas de investigación que merezcan la pena estudiarse. Una revisión sistemática de la investigación relativa a un campo de investigación de interés es un buen lugar para comenzar. La asistencia a conferencias, y estar alerta a los nuevos resultados, amplía la experiencia del investigador más allá de lo que ya se ha publicado.
- La decisión única más importante que toma un nuevo investigador es la elección de uno o dos científicos con experiencia para que sean sus mentores: investigadores experimentados que dedicarán tiempo a reunirse, ofrecerán recursos y contactos, estimularán la creatividad, y fomentarán la independencia y la visibilidad de los científicos jóvenes a su cargo.
- Las buenas preguntas de investigación surgen de encontrar buenos colaboradores en conferencias, del pensamiento crítico sobre la práctica clínica y sus problemas, de la aplicación de nuevos métodos a viejos problemas, y de considerar ideas que surgen de la docencia, de las ensoñaciones diurnas y de la búsqueda tenaz de soluciones a problemas graves.
- Antes de dedicar mucho tiempo y esfuerzo a escribir una propuesta o a llevar a cabo un estudio, el investigador debe tener en cuenta si la pregunta y el plan del estudio son: factibles, interesantes, novedosos, éticos y relevantes. Quienes financian investigaciones dan prioridad a las propuestas que puedan tener consecuencias innovadoras y significativas en la ciencia y la salud.
- Al principio, la pregunta de investigación debe desarrollarse en un resumen del estudio escrito en una página, que describa específicamente cuántos participantes serán necesarios, y cómo se elegirán estos participantes y se realizarán las mediciones.
- El desarrollo de la pregunta de la investigación y el plan del estudio es un proceso iterativo que incluye consultas con asesores y amigos, una creciente familiaridad con la bibliografía, y estudios piloto de los abordajes del reclutamiento y de las mediciones.
- La mayor parte de los estudios tienen más de una pregunta, aunque es útil centrarse en una única pregunta principal al diseñar y realizar el estudio.
- La investigación aplicada es un tipo de investigación clínica que estudia la aplicación de los hallazgos básicos a los estudios clínicos de pacientes (T1), así como el modo de aplicar estos hallazgos para mejorar las prácticas sanitarias en la comunidad (T2); necesita la colaboración entre investigadores de laboratorio e investigadores basados en la población, usando los métodos de investigación clínica que se presentan en este libro.

Volver al principio

BIBLIOGRAFÍA

1. The ATAC Trialists Group. Anastrazole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early breast cancer: first results of the ATAC randomized trials. Lancet 2002;359:2131-2139. Citado aquí

2. Quinn J, Cummings S, Callaham M, et al. Suturing versus conservative management of lacerations of the hand: randomized controlled trial. BMJ 2002;325:299-301. Citado aquí

3. Kuhn TS. The structure of scientific revolutions. Chicago, IL: University of Chicago Press, 1962. Citado aquí

4. Yaffe K, Browner W, Cauley J, et al. Association between bone mineral density and cognitive decline in older women. J Am Geriatr Soc 1999;47:1176-1182. Citado aquí

5. Prentice RL, Caan B, Chlebowski RT, et al. Low-fat dietary pattern and risk of invasive breast cancer. JAMA 2006;295:629-642. Citado aquí

6. Zerhouni EA. US biomedical research: basic, translational and clinical sciences. JAMA 2005;294:1352-1358. Citado aquí

CAPÍTULO 3 Elección de los participantes del estudio: especificación, muestreo e inclusión

Stephen B. Hulley
Thomas B. Newman
Steven R. Cummings

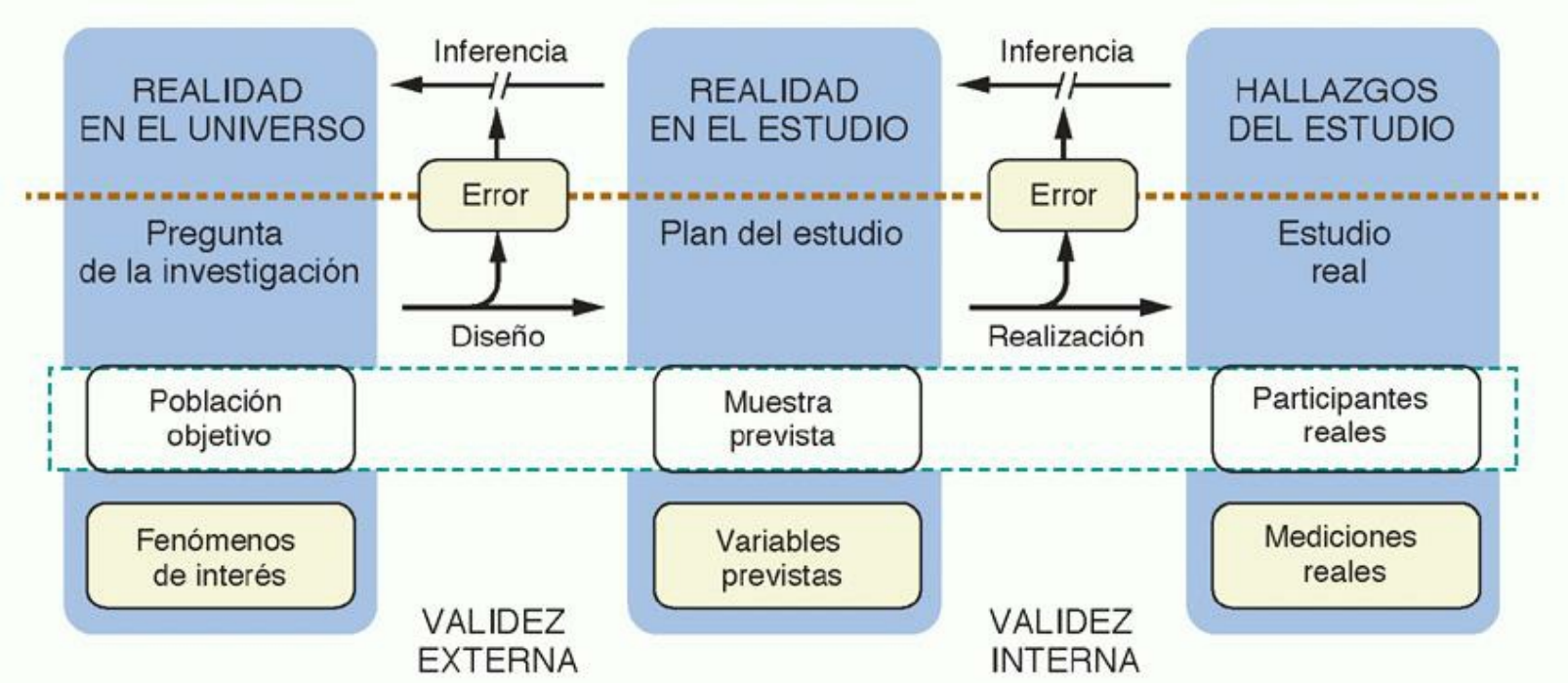
Una buena elección de los participantes del estudio cumple el objetivo esencial de asegurar que sus hallazgos representen de forma exacta lo que sucede en la población de interés. El protocolo debe especificar una muestra de participantes que pueda estudiarse con un coste de tiempo y económico aceptable (es decir, de tamaño moderado y de acceso sencillo), pero que sea lo suficientemente grande para controlar el error aleatorio y lo suficientemente representativa para permitir la generalización de los hallazgos del estudio a las poblaciones de interés. Un importante precepto aquí es que la posibilidad de generalizar rara vez es un sencillo asunto de sí o no; es una opinión cualitativa compleja que depende de la elección de la población y del diseño del muestreo por parte del investigador.

Regresaremos al tema de la elección del *número* adecuado de participantes del estudio en el capítulo 6. En este capítulo se orienta el proceso de especificación y muestreo de los *tipos* de participantes que serán representativos y viables (fig. 3-1). También se comentan las estrategias para la selección de estas personas que van a participar en el estudio.

▪ TÉRMINOS Y CONCEPTOS BÁSICOS

Poblaciones y muestras

Una población es un grupo completo de personas con unas características especificadas, y una muestra es un subgrupo de la población. En el uso no profesional, las características que definen una población tienden a ser geográficas: por ejemplo, la población de Canadá. En investigación, las características que la definen son también clínicas, demográficas y temporales:



▪FIGURA 3-1. Este capítulo aborda la elección de una muestra de participantes en el estudio que represente a la población de interés para la pregunta de la investigación.
Herramientas de imágenes

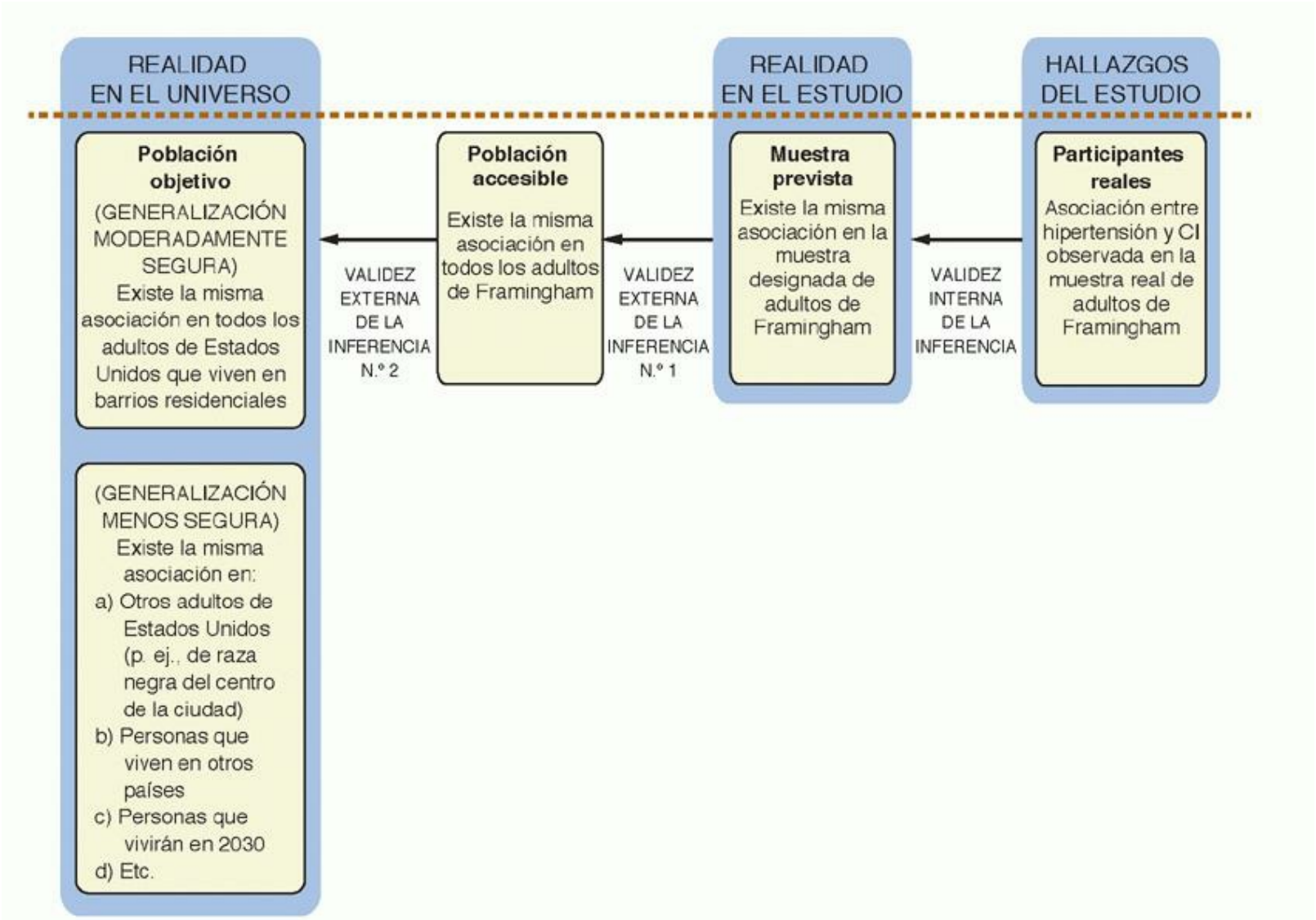
- Las características clínicas y demográficas definen la población objetivo, el gran grupo de personas de todo el mundo a las que se pueden generalizar los resultados: los adolescentes con asma, por ejemplo.
- La población accesible es un subgrupo de la población objetivo definido geográfica y temporalmente, y del que puede disponerse para su estudio: adolescentes con asma que viven este año en la ciudad del investigador.
- La muestra del estudio prevista es el subgrupo de población accesible que el investigador intenta incluir en el estudio.
- La muestra real del estudio es el grupo real de personas que participan en el estudio.

Generalización de los hallazgos del estudio

El clásico estudio de Framingham fue uno de los primeros abordajes del diseño científico de un estudio para poder aplicar a una población las inferencias obtenidas a partir de los hallazgos observados en una muestra (fig. 3-2).

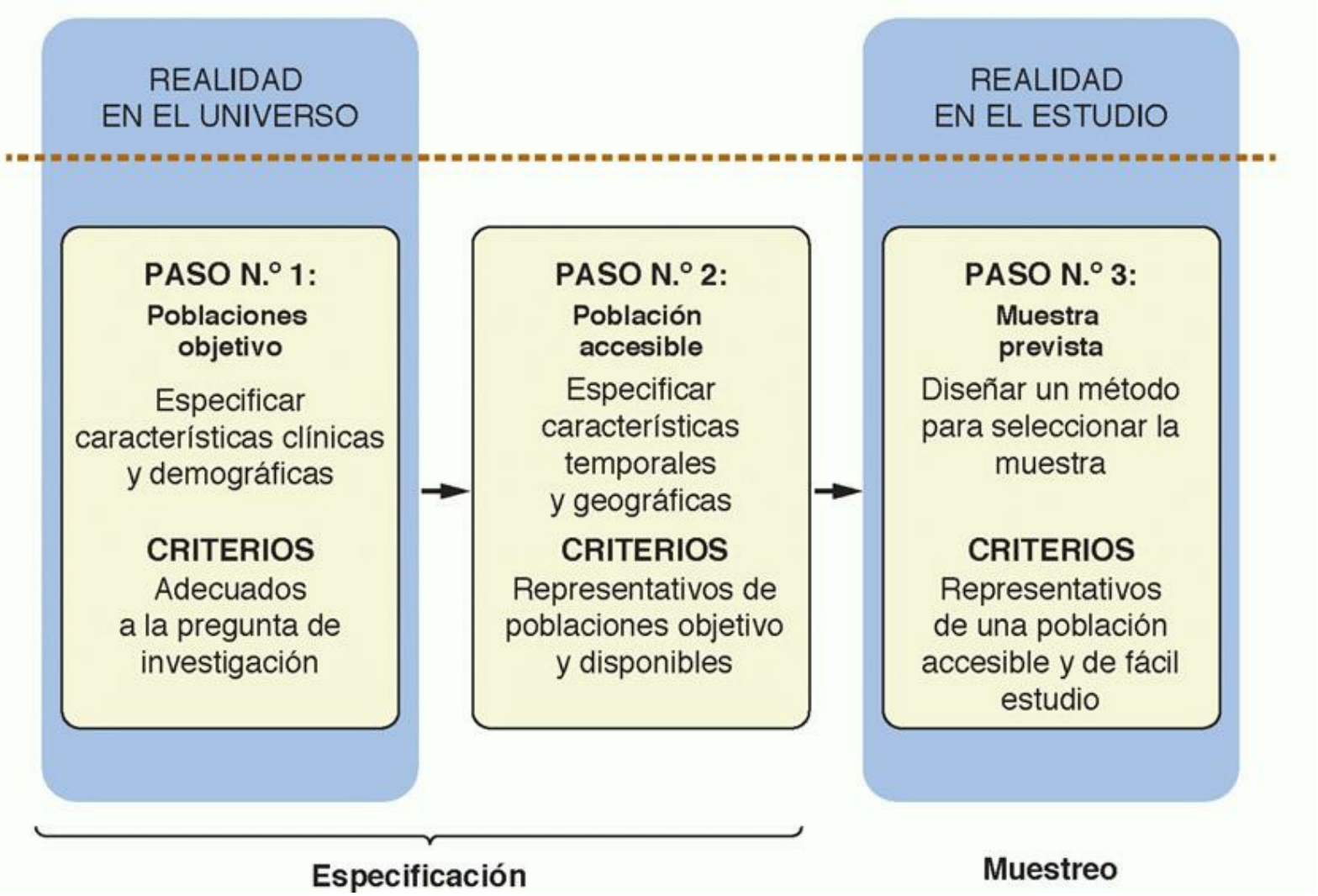
El diseño del muestreo exigía la identificación de todas las familias de Framingham con al menos una persona de 30-59 años, la elaboración de una lista de las familias por orden de dirección, y después la petición de participación a todas las personas en edad de participar de las dos primeras de cada conjunto de tres familias. Este diseño «sistemático» del muestreo no es tan a prueba de manipulaciones como elegir a cada uno de los participantes mediante un proceso aleatorio (como se discute más adelante en este capítulo), pero dos problemas más graves fueron que un tercio de los residentes en Framingham a los que se había seleccionado para el estudio rechazó participar, y que en su lugar los investigadores aceptaron a residentes en edad de participar que no estaban en la muestra y que se presentaron voluntarios (1).

Debido a que los que dan una respuesta favorable suelen estar más sanos que los que no lo hacen, especialmente si son voluntarios, las características de la muestra real indudablemente diferían de las de la muestra prevista. Todas las muestras tienen algunos errores, sin embargo, y el asunto es hasta qué punto se ha producido un daño. Los errores de muestreo del estudio de Framingham no parecen ser lo suficientemente importantes como para invalidar la conclusión de que sus hallazgos (p. ej., que la hipertensión es un factor de riesgo de sufrir cardiopatía isquémica [CI]) pueden generalizarse a todos los residentes de Framingham.



▪FIGURA 3-2. Las inferencias en la generalización desde los participantes del estudio a las poblaciones objetivo avanzan de derecha a izquierda.

Herramientas de imágenes



▪FIGURA 3-3. Pasos del diseño del protocolo para elegir los participantes del estudio.
Herramientas de imágenes

El segundo problema es la validez que tiene la generalización del hallazgo de que la hipertensión es un factor de riesgo de CI obtenido en la población accesible de los residentes de Framingham a poblaciones objetivo de cualquier otro lugar. Esta inferencia es más subjetiva. Se escogió la ciudad de Framingham no con un diseño científico del muestreo, sino porque parecía bastante típica de las comunidades residenciales de clase media de Estados Unidos y a los investigadores les venía bien. La validez de la generalización de las relaciones de riesgo de Framingham a poblaciones de otros puntos del país conlleva la norma de que, en general, los ensayos clínicos y estudios analíticos que abordan relaciones biológicas producen resultados más ampliamente generalizables en las diversas poblaciones que los estudios descriptivos que abordan distribuciones de características. Así, la importancia de la hipertensión como factor de riesgo de CI es similar en los residentes caucásicos de Framingham y en los afroamericanos de ciudades del interior, pero la frecuencia de la hipertensión es mucho mayor en esta última población.

Pasos del diseño del protocolo para seleccionar participantes para el estudio

Las inferencias de la figura 3-2 se presentan de derecha a izquierda, la secuencia que se usa para interpretar los hallazgos de un estudio completo. Un investigador que está planificando un estudio invierte esta secuencia, empezando por la izquierda (fig. 3-3). Comienza especificando las características clínicas y demográficas de la población objetivo que sirven para la pregunta de la investigación. A continuación utiliza criterios geográficos y temporales para especificar una muestra del estudio que sea representativa y práctica.

Volver al principio

▪ CRITERIOS DE SELECCIÓN

Un investigador que desea comparar la eficacia de los suplementos de testosterona en dosis baja con un placebo para aumentar la libido en mujeres posmenopáusicas empieza creando criterios de selección que definan la población que se va a estudiar.

Establecimiento de criterios de selección

Los criterios de inclusión definen las principales características de la población objetivo que corresponden a la pregunta de la investigación (tabla 3-1). La edad es, a menudo, un factor crucial, y en este estudio el investigador podría decidir centrarse en mujeres de 50 a 60 años de edad, especulando que en este grupo la relación entre el beneficio y el perjuicio del fármaco podría ser óptima; otro estudio podría tomar una decisión diferente y estudiar a mujeres de más edad. El investigador también podría incorporar al estudio a mujeres afroamericanas, hispanas y asiáticas, en un intento de ampliar la posibilidad de generalización. Generalmente esto es una buena idea, pero es necesario comprender que el aumento de la posibilidad de generalización es ilusorio si hay otros datos que indiquen que los efectos difieren de unas razas a otras. En este caso, el investigador debería tener suficientes mujeres de cada raza para comprobar estadísticamente la presencia de modificación del efecto (un efecto en una raza es diferente del que se ve en otras razas, lo que también se conoce como «interacción»; cap. 9); el número necesario generalmente es grande, y la mayor parte de los estudios no tienen potencia para detectar modificaciones del efecto.

Los criterios de inclusión relativos a las características geográficas y temporales de la población accesible muchas veces suponen concesiones entre objetivos científicos y objetivos prácticos. El investigador puede encontrar que los pacientes de su propio hospital son una fuente disponible y económica de pacientes, aunque debe preguntarse si las peculiaridades de los patrones de derivación locales podrían interferir con la generalización de los resultados a otras poblaciones. Sobre estas y otras decisiones en relación con los criterios de inclusión no hay ninguna acción única que sea claramente correcta o incorrecta; lo importante es tomar decisiones que sean sensatas, que se puedan utilizar de manera constante durante todo el estudio y que se puedan describir claramente a otras personas que decidan a quién se aplican las conclusiones publicadas.

Especificar características clínicas para seleccionar a los participantes a menudo implica opiniones difíciles, no solo sobre qué factores son importantes para la pregunta de la investigación, sino sobre cómo definirlos. ¿Cómo, por ejemplo, pondría un investigador en práctica el criterio de que los participantes tengan un «buen estado general de salud»? Podría decidir no incluir a pacientes con enfermedades referidas por ellos mismos, aunque probablemente esto excluyera a un gran número de pacientes que son perfectamente adecuados para la pregunta de la investigación que se quiere abordar.

	CARACTERÍSTICA DEL DISEÑO	**EJEMPLO**
Criterios de inclusión (sea específico)	Especificar poblaciones pertinentes a la pregunta de la investigación y eficientes para el estudio:	
	Características demográficas	Mujeres de 50 a 60 años de edad
	Características clínicas	Buena salud general
		Con pareja sexual
		Preocupado por el descenso de la libido
	Características geográficas (administrativas)	Pacientes que acuden a la consulta en el hospital del investigador
	Características temporales	Entre el 1 de enero y el 31 de diciembre del año especificado
Criterios de exclusión (sea moderado)	Especificar personas de la población que no se estudiarán debido a:	
	Una gran probabilidad de que se pierdan durante el	Alcohólica

	seguimiento	Tiene previsto mudarse a otra ciudad
	Imposibilidad de obtener buenos datos	Desorientada Tiene una barrera idiomática*
	Presentar un riesgo elevado de posibles efectos adversos	Antecedente de infarto de miocardio o accidente cerebrovascular

* Las alternativas a la exclusión de participantes con barreras idiomáticas (cuando estos subgrupos tienen un gran tamaño y son importantes para la pregunta de la investigación) serían recopilar datos no verbales o utilizar personal y cuestionarios bilingües.

Sería más razonable que pudiera excluir solo a los que tengan enfermedades que pudieran interferir con el seguimiento, como cáncer metastásico. Esto sería un ejemplo de «criterios de exclusión», que se refieren a las personas que cumplen los criterios de inclusión y que serían adecuadas para el estudio de no ser por características que podrían interferir con el éxito de los intentos de seguimiento, la calidad de los datos o la aceptabilidad del tratamiento aleatorizado (v. tabla 3-1). La dificultad con el idioma español, los problemas psicológicos, el alcoholismo y las enfermedades graves son ejemplos de criterios de exclusión. Los ensayos clínicos difieren de los estudios de observación en que es más probable que haya exclusiones exigidas por la preocupación de la seguridad de una intervención en determinados pacientes; por ejemplo, el uso de fármacos en embarazadas (cap. 10). Una buena regla general que hace que las cosas sigan siendo sencillas y conserva el número de posibles participantes en el estudio es tener el *menor número de criterios de exclusión* posible.

Poblaciones clínicas y de la comunidad

Si la pregunta de la investigación afecta a pacientes con una enfermedad, es más fácil encontrar pacientes hospitalizados o en consultas, pero los factores de selección que determinan quién acude al hospital o a la consulta pueden tener un efecto importante. Por ejemplo, una consulta especializada en un centro médico de atención terciaria atrae a pacientes con formas graves de enfermedad, lo que da una impresión distorsionada de las características y el pronóstico que se ven en la práctica habitual. El muestreo en consultas de atención primaria sería una opción mejor.

Otra opción frecuente al elegir la muestra es seleccionar participantes de la comunidad que representen una población sana. Estas muestras se reúnen a veces utilizando el correo, el correo electrónico o anuncios en Internet, emisoras de radio o medios impresos; no son totalmente representativas de una población general, porque algunos tipos de personas tienen más probabilidad que otras de participar voluntariamente o de ser usuarios activos de Internet o de correo electrónico. Las verdaderas muestras «basadas en la población» son difíciles y caras de reunir, pero son útiles para guiar las políticas de salud pública y el ejercicio médico en la comunidad. Uno de los mayores y mejores ejemplos es el estudio *National Health and Nutrition Survey* (NHANES), una muestra representativa de todos los residentes de Estados Unidos.

Puede aumentarse el tamaño y la diversidad de la muestra colaborando con compañeros de otras ciudades, o usando series de datos ya existentes, como los datos del estudio NHANES y de Medicare. Las series de datos de acceso electrónico de agencias de salud pública, organizaciones de prestación de asistencia sanitaria y compañías de seguro médico se usan cada vez más en investigación clínica, y pueden ser más representativas de poblaciones nacionales y requerir menos tiempo que otras posibilidades (cap. 13).

Volver al principio

▪ MUESTREO

A menudo, el número de personas que cumplen los criterios de selección es demasiado grande, y es necesario seleccionar una muestra (subgrupo) de la población para su estudio.

Muestras no probabilísticas

En investigación clínica, la muestra del estudio está formada a menudo por personas que cumplen los criterios de admisión y a las que el investigador tiene fácil acceso. Es lo que se denomina una muestra de conveniencia, y tiene ventajas evidentes en cuanto a coste y

logística, lo que hace que sea una buena elección para algunas preguntas de investigación.

Una muestra consecutiva puede reducir al mínimo el sesgo de participación voluntaria y otros sesgos de selección, al elegir de forma consecutiva a los pacientes que cumplan los criterios de admisión. Este abordaje es especialmente deseable, por ejemplo, cuando significa incluir a toda la población accesible a lo largo de un período lo suficientemente largo como para que puedan producirse variaciones estacionales u otros cambios temporales que son importantes para la pregunta de la investigación.

La validez de extraer inferencias de cualquier muestra es la premisa de que, con el fin de responder la pregunta de la investigación, representa suficientemente la población. Con muestras de conveniencia, esto requiere una opinión subjetiva.

Muestras probabilísticas

A veces, particularmente con preguntas de investigación descriptivas, es necesaria una base científica para generalizar los hallazgos de la muestra en estudio a la población. El muestreo probabilístico, el método de referencia para asegurar la posibilidad de generalizar los resultados, utiliza un proceso aleatorio para garantizar que cada unidad de la población tenga una posibilidad específica de verse incluida en la muestra. Se trata de un método científico que proporciona una base rigurosa para calcular la fidelidad con que los fenómenos observados en la muestra representan los de la población, y para calcular la significación estadística y los intervalos de confianza. Existen varias versiones de este método:

- Una muestra aleatoria simple se extrae enumerando (contando) todas las personas de la población de la que se extraerá la muestra, y seleccionando aleatoriamente un subgrupo. El uso más frecuente de este método en investigación clínica es cuando el investigador desea seleccionar un subgrupo representativo de una población que es mayor de lo que necesita. Para tomar una muestra aleatoria de los pacientes operados de cataratas en su hospital, por ejemplo, el investigador podría obtener una lista de todos estos pacientes de los partes de quirófano durante el período del estudio, y usar después una tabla de números aleatorios para seleccionar pacientes para su estudio (apéndice 3).

- Una muestra sistemática es similar a una muestra aleatoria simple en el primer paso, la enumeración de la población, pero difiere en que la muestra se elige mediante un proceso periódico predeterminado (p. ej., el método Framingham de obtener las primeras dos de cada tres familias de una lista de familias residentes en la ciudad ordenadas por dirección). El muestreo sistemático puede presentar errores causados por periodicidades naturales en la población, y permite al investigador predecir, y quizás manipular, quiénes estarán en la muestra. No ofrece ventajas logísticas sobre el muestreo aleatorio simple, y rara vez es una opción mejor en la investigación clínica.

- Una muestra aleatoria estratificada comienza dividiendo la población en subgrupos según características como el sexo o la raza, y se toma una muestra aleatoria de cada uno de esos «estratos». Pueden asignarse pesos a las submuestras de una muestra estratificada para extraer una proporción anormalmente elevada de subgrupos que son menos frecuentes en la población, pero que tienen un interés especial para el investigador. Al estudiar la incidencia de la toxemia en el embarazo, por ejemplo, el investigador podría estratificar la población según la raza y, a continuación, obtener muestras de igual tamaño de cada estrato. Entonces, las razas menos frecuentes estarían sobrerrepresentadas, lo que permitiría obtener estimaciones de la incidencia con una precisión comparable en todos los grupos raciales.

- Una muestra de conglomerados es una muestra aleatoria de agrupaciones (conglomerados) naturales de individuos de la población. Este muestreo es muy útil cuando la población está muy dispersa y no se puede realizar una lista y hacer un muestreo de todos sus elementos. Considere, por ejemplo, el problema de entrevistar a pacientes con cáncer de pulmón seleccionados aleatoriamente de una base de datos de diagnósticos al alta hospitalaria de ámbito estatal; se podría estudiar a los pacientes con un coste menor escogiendo una muestra aleatoria de los hospitales y obteniendo los casos de estos. Los estudios en la comunidad a menudo usan una muestra de conglomerados en dos etapas: una muestra aleatoria de manzanas de la ciudad se obtiene de las manzanas de la ciudad enumeradas en un mapa, y un equipo de campo visita las manzanas de la muestra, hace una lista de todas las direcciones de cada una y selecciona una submuestra para su estudio mediante un segundo proceso aleatorio. Un inconveniente del muestreo de conglomerados es que los grupos que aparecen de forma natural son, a menudo, más homogéneos en relación con las variables de interés que la población; cada manzana de la ciudad, por ejemplo, tiende a albergar personas de un nivel socioeconómico similar. Esto significa que el tamaño eficaz de la muestra (después de ajustar la uniformidad dentro de los conglomerados) será algo menor que el número de participantes, y que el análisis estadístico debe tener en cuenta el agrupamiento.

Resumen de las opciones de diseño del muestreo

El uso de estadísticos descriptivos y pruebas de significación estadística para extraer inferencias sobre la población a partir de observaciones en la muestra del estudio se basa en la suposición de que se ha utilizado una muestra probabilística. En investigación

clínica, sin embargo, casi nunca es posible tener una muestra aleatoria de toda la población objetivo. El muestreo de conveniencia, preferiblemente con un diseño consecutivo, es un enfoque práctico que a menudo resulta adecuado. La decisión sobre si el diseño de muestreo propuesto es satisfactorio necesita que el investigador emita un juicio: *para la pregunta de la investigación que se plantea,* ¿las conclusiones que se extraigan de las observaciones del estudio serán similares a las conclusiones que se obtendrían del estudio de una muestra probabilística verdadera de la población accesible? Y, aparte de eso, ¿las conclusiones serán adecuadas para la población objetivo?

Volver al principio

▪ SELECCIÓN

Objetivos de la selección

Un factor importante que hay que tener en cuenta en el momento de elegir la población accesible y el método de muestreo es la posibilidad de reunir participantes para el estudio. Dos son los objetivos principales: *1)* reunir una muestra que represente adecuadamente a la población objetivo, minimizando la posibilidad de obtener una respuesta incorrecta a la pregunta de la investigación debido al error sistemático (sesgo), y *2)* reunir un tamaño de la muestra suficiente para minimizar la probabilidad de obtener una respuesta incorrecta debido al error aleatorio (por el azar).

Obtención de una muestra representativa

El abordaje para conseguir una muestra representativa empieza en la fase de diseño con decisiones sensatas sobre la elección de poblaciones objetivo accesibles y los métodos de muestreo. Termina con la ejecución, protegiéndose frente a los errores en la aplicación de los criterios de selección de los participantes prospectivos del estudio, y reforzando las estrategias eficaces a medida que avance el estudio. Un problema importante, especialmente en los estudios de observación, es el problema de la falta de respuesta [1] . La proporción de pacientes seleccionados para el estudio que dan su consentimiento a ser incluidos (la tasa de respuesta) influye en la validez de la inferencia de que la muestra incluida representa a la población. Las personas a las que es difícil llegar, y las que rechazan participar una vez que se ha establecido contacto con ellas, tienden a ser diferentes de las personas que sí se incorporan. El nivel de falta de respuesta que pondrá en peligro la posibilidad de generalización del estudio depende de la naturaleza de la pregunta de la investigación y de los motivos para no responder. Una tasa de falta de respuesta del 2 %, que sería un buen logro en muchas situaciones, puede distorsionar gravemente la prevalencia estimada de una enfermedad cuando la propia enfermedad es una causa de falta de respuesta.

El grado en que el sesgo por falta de respuesta puede influir en las conclusiones de un estudio descriptivo a veces se puede estimar durante el estudio, obteniendo información adicional de una muestra de personas que no responden. Sin embargo, la mejor forma de abordar el sesgo de falta de respuesta es reducir al mínimo el número de personas que no responden. El problema de la imposibilidad de contactar con personas a las que se ha escogido para la muestra puede reducirse diseñando una serie sistemática de intentos de contacto repetidos usando varios métodos (correo postal, correo electrónico, teléfono, visita domiciliaria). Puede reducirse al mínimo el rechazo a participar en personas con las que se contacta mejorando la eficacia y el atractivo del estudio, escogiendo un diseño que evite pruebas cruentas y desagradables, usando folletos y conversaciones individuales para aliviar la ansiedad y el malestar, proporcionando incentivos como el reembolso del coste del transporte y la entrega de los resultados de las pruebas, y evitando las barreras idiomáticas con personal bilingüe y cuestionarios traducidos.

Selección de cantidades suficientes de participantes

Quedarse corto en la tasa de selección es uno de los problemas más habituales en investigación clínica. Al planificar un estudio, es mejor suponer que la cantidad de personas que cumplirán los criterios de inclusión y que estarán de acuerdo en participar en el estudio será menor (a menudo, varias veces menor) que el número proyectado al principio. Los abordajes para resolver este problema son: calcu lar la magnitud del problema de selección empíricamente mediante una prueba previa, planificar el estudio con una población accesible que sea mayor de lo que se cree que será necesario, y elaborar planes de contingencia por si surge la necesidad de obtener participantes adicionales. Mientras se realiza la selección, es importante controlar estrechamente el progreso en cuanto al cumplimiento de los objetivos de la selección y tabular las razones por las que no se llega a los objetivos. Conocer por qué se pierde a los posibles participantes en el estudio en diferentes fases puede llevar a estrategias para reducir estas pérdidas. A veces, la inclusión supone seleccionar pacientes que los miembros del equipo de investigación ya conocen (p. ej., en un estudio de un nuevo tratamiento en pacientes que acuden a la consulta del investigador). Aquí, el principal problema es presentar de manera justa la oportunidad de participar en el estudio, dejando claros las ventajas y los inconvenientes. Al comentar la participación, el investigador debe reconocer los dilemas éticos que surgen cuando su consejo, como médico del paciente, podría entrar en conflicto con sus intereses como investigador (cap. 14).

A menudo, la selección conlleva entrar en contacto con poblaciones desconocidas para los miembros del equipo investigador. Será útil que al menos un miembro del equipo tenga alguna experiencia con los métodos para contactar con los posibles participantes. Entre ellos se encuentran: detección selectiva en entornos laborales o lugares públicos, como centros comerciales; envío de gran cantidad de publicidad por correo a listados de, por ejemplo, personas con carné de conducir; publicidad en Internet; invitación a remisiones por parte de otros médicos; revisión retrospectiva de historias clínicas; y examen de listas de pacientes atendidos en consultas y hospitales. Algunos de estos métodos, en concreto los dos últimos, acarrean temas como la intromisión en la vida privada, que debe tener en cuenta el comité de ética de la investigación médica.

Puede resultar útil prepararse para la selección obteniendo el apoyo de organizaciones importantes. Por ejemplo, el investigador puede reunirse con administradores del hospital para hablar sobre una muestra de pacientes de la consulta, y con los dirigentes comunitarios, la sociedad médica y el departamento de salud de la región para planificar una operación de detección en la comunidad o el envío de cartas a los médicos. Pueden incluirse apoyos por escrito, como un apéndice en las solicitudes de financiación. En estudios de gran tamaño puede ser útil la creación de un clima favorable en la comunidad mediante conferencias públicas, o con publicidad en la radio, la televisión, la prensa, folletos, páginas web y envíos publicitarios masivos por correo.

Volver al principio

▪ RESUMEN

- La mayor parte de la investigación clínica se basa, desde el punto de vista filosófico y práctico, en el uso de una muestra que represente a una población.
- La ventaja del muestreo es la eficiencia; permite al investigador extraer inferencias sobre una población de gran tamaño examinando un subgrupo con un coste relativamente pequeño en cuanto a tiempo y esfuerzo. El inconveniente está en las fuentes de error que introduce. Si la muestra no es suficientemente representativa para la pregunta de la investigación, los hallazgos pueden no generalizarse bien a la población objetivo, y, si no es suficientemente grande, los hallazgos pueden no minimizar la importancia del azar.
- Cuando diseña una muestra, el investigador comienza conceptualizando la población objetivo con un grupo específico de participantes que se adapten bien a la pregunta de la investigación.
- Después, selecciona una población accesible y adecuada, que sea asequible desde los puntos de vista geográfico y temporal, y define un conjunto económico de criterios de exclusión que eliminen a las personas cuyo estudio no sería ético o adecuado.
- El siguiente paso será diseñar un método de muestreo en la población. Puede ser adecuada una muestra de conveniencia, especialmente para el estudio inicial de algunas preguntas, y muchas veces es una buena opción una muestra consecutiva. Puede realizarse un muestreo aleatorio simple para reducir el tamaño de una muestra de conveniencia, si es necesario; en determinadas situaciones son útiles otras estrategias de muestreo probabilístico (estratificado y por conglomerados).
- Finalmente, el investigador debe diseñar y llevar a cabo estrategias para la selección de una muestra de participantes que sea suficientemente representativa de la población objetivo para controlar las fuentes sistemáticas de error, y suficientemente grande para controlar las fuentes aleatorias de error.

Volver al principio

[1]La preocupación por la ausencia de respuesta en el proceso de *selección* de los pacientes para un estudio (el tema de este capítulo) es principalmente una preocupación en estudios descriptivos que tienen como objetivo principal estimar la distribución de variables en poblaciones particulares. La ausencia de respuesta en el proceso de *seguimiento* es muchas veces un problema importante en cualquier estudio en el que se siga a una cohorte a lo largo del tiempo, y particularmente en un ensayo clínico de una intervención que puede alterar la tasa de respuesta (cap. 10).

APÉNDICE 3
NA

Esta tabla presenta una sencilla forma en papel para seleccionar una muestra aleatoria del 10 % gracias a una tabla de números aleatorios. Comience enumerando (realizando una lista y asignando un número) a todas las personas de la población de la que se va a

obtener la muestra. Después, decida una regla para obtener una serie adecuada de números; por ejemplo, si su lista tiene 741 elementos (a los que se han asignado números del 1 a 741), la regla podría ser recorrer verticalmente hacia abajo cada una de las columnas de esta tabla, utilizando los primeros tres dígitos de cada número (comenzando en la esquina superior izquierda, los números son 104, 223, etc.), y seleccionar los primeros 74 números diferentes que se encuentren en el intervalo de 1 a 741. Finalmente, seleccione un punto de partida mediante un proceso arbitrario (cerrar los ojos y poner el lápiz en algún número de la tabla es una forma de hacerlo) y comience a aplicar la regla. El abordaje moderno, con una serie computarizada de números aleatorios, funciona básicamente de la misma manera.

10480	15011	01536	81647	91646	02011
22368	46573	25595	85393	30995	89198
24130	48390	22527	97265	78393	64809
42167	93093	06243	61680	07856	16376
37570	33997	81837	16656	06121	91782
77921	06907	11008	42751	27756	53498
99562	72905	56420	69994	98872	31016
96301	91977	05463	07972	18876	20922
89572	14342	63661	10281	17453	18103
85475	36857	53342	53998	53060	59533
28918	79578	88231	33276	70997	79936
63553	40961	48235	03427	49626	69445
09429	93969	52636	92737	88974	33488
10365	61129	87529	85689	48237	52267
07119	97336	71048	08178	77233	13916
51085	12765	51821	51259	77452	16308
02368	21382	52404	60268	89368	19885

01011 54092 33362 94904 31273 04146

52162 53916 46369 58569 23216 14513

07056 97628 33787 09998 42698 06691

48663 91245 85828 14346 09172 30163

54164 58492 22421 74103 47070 25306

32639 32363 05597 24200 38005 13363

29334 27001 87637 87308 58731 00256

02488 33062 28834 07351 19731 92420

81525 72295 04839 96423 24878 82651

29676 20591 68086 26432 46901 20949

00742 57392 39064 66432 84673 40027

05366 04213 25669 26422 44407 44048

91921 26418 64117 94305 26766 25940

BIBLIOGRAFÍA

1. www.framinghamheartstudy.org/about/background.html, último acceso, 7/23/12. Citado aquí

CAPÍTULO 4 Planificación de las mediciones: precisión, exactitud y validez

Stephen B. Hulley
Thomas B. Newman
Steven R. Cummings

Las mediciones describen fenómenos en términos que pueden analizarse estadísticamente, y la validez de un estudio depende de hasta qué punto las variables diseñadas para el estudio representan los fenómenos de interés (fig. 4-1). Por ejemplo, ¿cómo mide de bien la glucosa sanguínea un glucómetro portátil, o cómo detecta la cantidad y la calidad del sueño un cuestionario de insomnio?

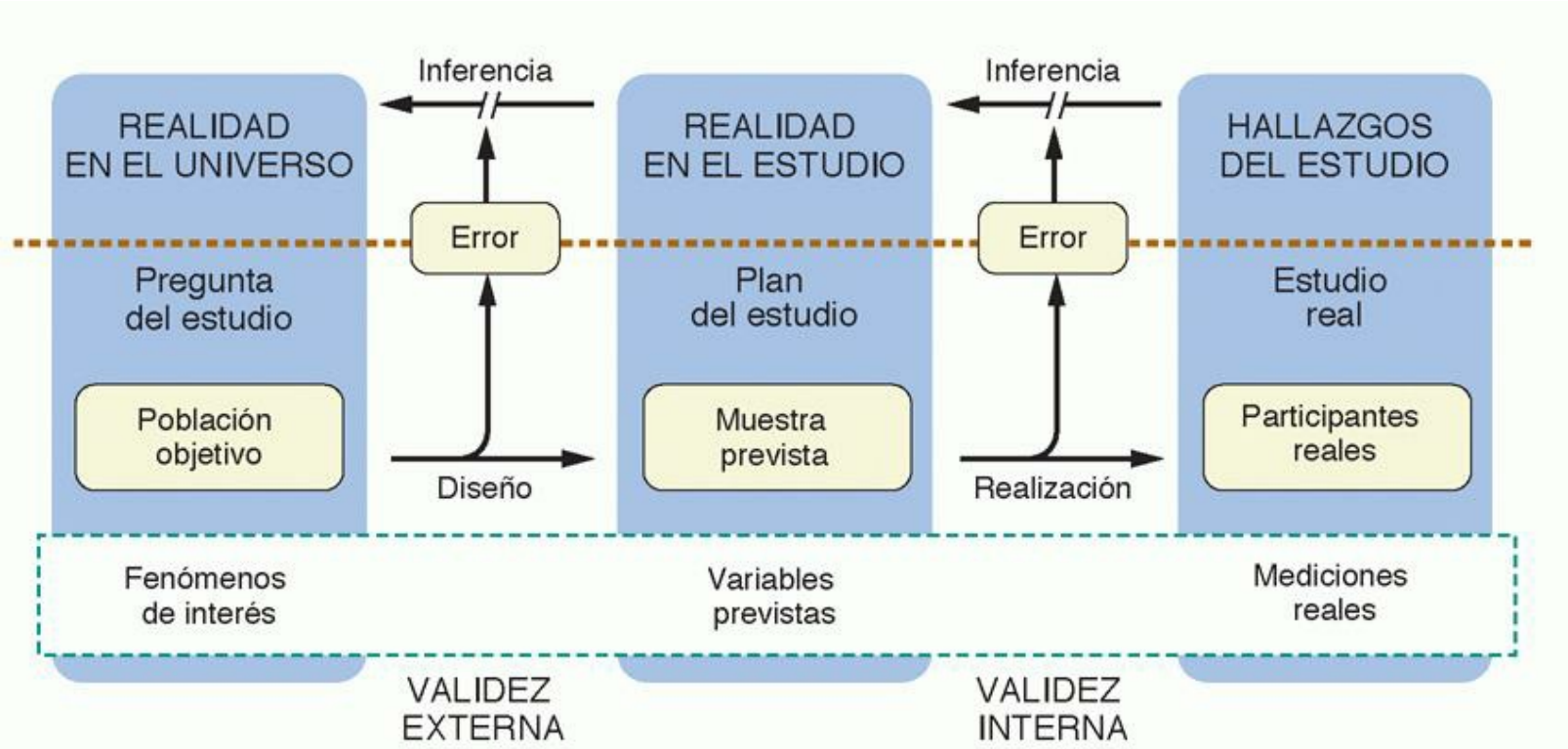
Este capítulo empieza considerando cómo la elección de una escala de medida influye en el contenido de la información de la medición. Vuelve después al objetivo central de reducir al mínimo el error de medición: cómo diseñar medidas que sean relativamente precisas (sin error aleatorio) y exactas (sin error sistemático), con lo que mejora la adecuación de la extracción de inferencias desde estas mediciones hasta los fenómenos de interés. Abordamos el concepto de validez, un familiar cualitativo de la exactitud, antes de concluir con algunas consideraciones para la investigación clínica y aplicada, señalando especialmente las ventajas de guardar muestras para determinaciones posteriores.

▪ ESCALAS DE MEDIDA

En la tabla 4-1 , se presenta una clasificación simplificada de escalas de medida y la información que se obtiene. La clasificación es importante, porque algunos tipos de variables son más informativos que otros, añadiendo potencia o reduciendo los requisitos de tamaño de la muestra, y revelando patrones de distribución más detallados.

Variables numéricas: variables continuas y discretas

Las variables numéricas se pueden cuantificar con un número que expresa cuánto o cuántos. Las variables continuas cuantifican cuánto en una escala infinita; el número de valores posibles del peso corporal, por ejemplo, está limitado solo por la sensibilidad del aparato que se utilice para medirlo. Las variables continuas poseen mucha información. Las variables discretas cuantifican cuántos en una escala con unidades fijas, habitualmente enteros, como el número de veces que ha estado embarazada una mujer. Las variables discretas que tienen un elevado número de valores posibles pueden parecer variables continuas en los análisis estadísticos y ser equivalentes para el objetivo de diseñar mediciones.



▪FIGURA 4-1. Diseño de mediciones que representen los fenómenos de interés.
Herramientas de imágenes

TIPO DE MEDICIÓN	CARACTERÍSTICAS DE LA VARIABLE	EJEMPLO	ESTADÍSTICO DESCRIPTIVO	POTENCIA ESTADÍSTICA

Categórica				
Dicotómica	Dos categorías	Estado vital (vivo o muerto)	Recuentos, proporciones	Baja
Nominal	Categorías no ordenadas	Raza, tipo sanguíneo	Igual que antes	Baja
Ordinal	Categorías ordenadas con intervalos que no se pueden cuantificar	Grado de dolor, clase social	Además de lo anterior: medianas	Intermedia
Numérica				
Continua o discreta†	Espectro ordenado con intervalos cuantificables	Peso, número de cigarrillos/día	Además de lo anterior: medias, desviaciones típicas	Elevada

† Las variables continuas tienen un número infinito de valores (p. ej., peso), mientras que las variables numéricas discretas son más limitadas (p. ej., número de cigarrillos/día). Las variables discretas que tienen un gran número de valores posibles parecen variables continuas con fines prácticos de medida y análisis.

Variables categóricas: dicotómicas, nominales y ordinales

Los fenómenos que no son adecuados para su cuantificación se miden clasificándolos en categorías. Las variables categóricas con dos posibles valores (p. ej., muerto o vivo) se denominan dicotómicas. Las variables categóricas con más de dos categorías (policotómicas) pueden, además, caracterizarse según el tipo de información que contienen. De estas, las variables nominales tienen categorías que no están ordenadas; el grupo sanguíneo O, por ejemplo, no es ni más ni menos que el grupo B; las variables nominales tienden a tener un carácter cualitativo absoluto para que sean sencillas de medir. Las categorías de las variables ordinales tienen un orden, como dolor intenso, moderado y leve. La información adicional es una ventaja sobre las variables nominales, pero, debido a que las variables ordinales no especifican una diferencia numérica o uniforme entre una categoría y la siguiente, el contenido de la información es menor que el de las variables numéricas discretas o continuas.

Elección de una escala de medida

Una buena norma general es preferir las variables continuas a las categóricas cuando se pueda elegir, porque la información adicional que contienen mejora la eficiencia estadística. En un estudio en el que se comparan los efectos antihipertensivos de varios tratamientos, por ejemplo, la determinación de la presión arterial en milímetros de mercurio permite al investigador observar la magnitud del cambio en cada participante, mientras que la medición como hipertensos o normotensos limita la evaluación. La variable continua contiene más información, y el resultado es un estudio con más potencia o con menor tamaño de la muestra (cap. 6).

Las variables continuas también ofrecen más flexibilidad que las variables categóricas para ajustar los datos a la naturaleza de la variable o a la forma de la asociación, especialmente cuando la relación podría tener un patrón complejo. Por ejemplo, en un estudio de la relación de la vitamina D con diversos cánceres se debería medir la vitamina D como variable continua para poder detectar un posible patrón con forma de U, según el cual se observa mayor mortalidad en personas con concentraciones bajas o elevadas de vitamina D que en las que tienen concentraciones intermedias (1). En un estudio de factores predictivos de bajo peso al nacimiento se debería registrar el peso real al nacimiento, en lugar de indicar si es mayor o menor que el umbral convencional de 2 500 g; esto deja abiertas las opciones analíticas, como modificar el valor de corte que define el bajo peso o el desarrollo de una escala ordinal con varias categorías de peso al nacimiento (p. ej., > 2 500 g, 2 000-2 499 g, 1 500-1 999 g y < 1 500 g).

Igualmente, cuando existe la opción de diseñar el número de categorías de respuesta en una escala ordinal (como en una pregunta sobre preferencias alimentarias), a menudo resulta útil proporcionar media docena de categorías que oscilen desde *no gustar nada* a *gustar enormemente*. Los resultados pueden quedar más adelante colapsados en una dicotomía (no gustar y gustar), pero no lo contrario.

Muchas características, sobre todo síntomas como el dolor o aspectos de los hábitos de vida, son difíciles de describir con categorías o números. Pero estos fenómenos, con frecuencia, son importantes en el diagnóstico y las decisiones terapéuticas, y el intento de medirlos es una parte esencial del método científico de descripción y análisis. Esto se ilustra mediante el cuestionario *Short Form* (SF)-36, un cuestionario normalizado para evaluar la calidad de vida que permite obtener puntuaciones numéricas discretas (2). El proceso de clasificación y medición, si se realiza bien, puede aumentar la objetividad de nuestro conocimiento, reducir los sesgos y proporcionar un medio de comunicación.

Volver al principio

▪ PRECISIÓN

La precisión de una variable es el grado en que es reproducible, con casi el mismo valor cada vez que se mida. Una balanza puede medir el peso corporal con gran precisión, mientras que es más probable que una entrevista para medir la calidad de vida produzca valores que varíen de un observador o una ocasión a otro. La precisión tiene una gran influencia en la potencia de un estudio. Cuanto más precisa sea la medida, mayor será la potencia estadística con un tamaño de la muestra concreto para calcular valores medios y comprobar hipótesis (cap. 6).

La precisión (denominada también reproducibilidad, fiabilidad y consistencia) es una función del error aleatorio (variabilidad por el azar); cuanto mayor sea el error, menos precisa será la medida. Existen tres fuentes principales de error aleatorio al realizar medidas:

- Variabilidad del observador: se debe al observador, e incluye factores como escoger palabras en una entrevista o tener habilidad para usar un instrumento mecánico.
- Variabilidad del instrumento: se debe al instrumento, e incluye factores ambientales cambiantes (p. ej., temperatura), el desgaste de los componentes mecánicos, lotes diferentes de reactivos, etc.
- Variabilidad del participante: se debe a la variabilidad biológica intrínseca de los participantes en el estudio, que no se relaciona con las variables en estudio, como la variabilidad debida a la hora del día de las mediciones o el tiempo desde la última medicación.

Evaluación de la precisión

La precisión se valora como la reproducibilidad de mediciones repetidas, ya sea comparando mediciones realizadas por la misma persona (reproducibilidad intraobservador) o por personas diferentes (reproducibilidad entre observadores). Igualmente, puede evaluarse para un instrumento o entre diversos instrumentos. La reproducibilidad de las variables continuas se expresa, a menudo, como la desviación típica intraobservador o el coeficiente de variación (desviación típica intraobservador dividida por la media) [1] . Para las variables categóricas se utiliza a menudo la concordancia porcentual, el coeficiente de correlación intraclase y el estadístico kappa (3 , 4 , 5).

Estrategias para mejorar la precisión

Existen cinco métodos para reducir al mínimo el error aleatorio y aumentar la precisión de las determinaciones (tabla 4-2):

- Normalización de los métodos de medida. Todos los protocolos de estudio deberían incluir instrucciones específicas para hacer las mediciones (definiciones operativas). Entre ellas se encuentran direcciones escritas sobre cómo preparar el entorno y el participante, cómo realizar y registrar la entrevista, cómo calibrar el instrumento, etc. (apéndice 4). Esta serie de materiales, parte del manual operativo, es esencial en los estudios grandes y complejos, y recomendable en los más pequeños. Aun cuando solo exista un único observador, las directrices específicas por escrito para realizar cada una de las mediciones ayudarán a que su ejecución sea uniforme a lo largo del estudio y servirán de base para describir los métodos cuando se publiquen los resultados.
- Formación y certificación de los observadores. La formación mejorará la homogeneidad de las técnicas de medición, especialmente cuando intervienen varios observadores. A menudo es deseable diseñar una prueba formal del dominio de las técnicas especificadas en el manual de instrucciones y certificar que los observadores han logrado el nivel recomendado de rendimiento (cap. 17).
- Perfeccionamiento de los instrumentos. Los instrumentos mecánicos y electrónicos pueden diseñarse para reducir la variabilidad. Igualmente, los cuestionarios y las entrevistas pueden escribirse para aumentar la claridad y evitar posibles ambigüedades (cap. 15).
- Automatización de los instrumentos. Pueden eliminarse las variaciones en la forma en que los observadores humanos realizan las mediciones mediante dispositivos mecánicos automáticos y cuestionarios de autorrespuesta.

- Repetición. La influencia del error aleatorio de cualquier tipo se reduce repitiendo la medición, y usando la media de las dos o de más determinaciones. Con esta estrategia aumentará notablemente la precisión, siendo la principal limitación el coste añadido y las dificultades prácticas para repetir las determinaciones.

Para cada una de las mediciones del estudio, el investigador debe decidir cuánto hincapié debe hacer en cada una de estas estrategias. Esta decisión puede basarse en la importancia de la variable, en la magnitud del posible problema con la precisión, y en la viabilidad y el coste de la estrategia. En general, las dos primeras estrategias (normalización y formación) deben usarse siempre, y la quinta (repetición) es una opción que se garantiza que mejora la precisión, siempre que sea viable y asequible.

	ESTRATEGIA PARA REDUCIR EL ERROR ALEATORIO	ORIGEN DEL ERROR ALEATORIO	EJEMPLO DE ERROR ALEATORIO	EJEMPLO DE ESTRATEGIA PARA EVITAR EL ERROR
1.	Normalización de los métodos de medición en un manual de instrucciones	Observador	Variación de la medición de la presión arterial (PA) por una velocidad variable de desinflado del manguito (a veces demasiado deprisa)	Especificar que el manguito debe desinflarse a 2 mm Hg/s
		Participante	Variación en la PA debida a variaciones en la duración de estar sentado tranquilo antes de la medición	Especificar que el participante esté sentado en una sala tranquila durante 5 min antes de tomar la PA
2.	Formación y aprobación del observador	Observador	Variación en la PA por técnica variable del observador	Formar al observador en técnicas normalizadas
3.	Perfeccionamiento del instrumento	Instrumento y observador	Variación en la PA por funcionamiento inadecuado del manómetro	Comprar un nuevo manómetro de alta calidad
4.	Automatización del instrumento	Observador	Variación en la PA por variable técnica del observador	Usar aparatos automáticos para medir la PA
		Participante	Variación en la PA por reacción emocional del participante	Usar aparatos automáticos para medir la PA
5.	Repetición de la determinación	Observador, participante e instrumento	Todas las determinaciones y todas las fuentes de variación	Usar la media de dos o más determinaciones de la PA

	PRECISIÓN	EXACTITUD
Definición	Grado en que una variable tiene casi el mismo valor cuando se mide varias veces	Grado en que una variable se aproxima al valor verdadero
Mejor modo de evaluación	Comparación entre repetidas medidas	Comparación con un patrón de referencia
Utilidad para el		

estudio	Aumenta la potencia para detectar efectos	Aumenta la validez de las conclusiones
Amenazado por	Error aleatorio (azar) al que contribuyen	Error sistemático (sesgo) al que contribuyen
	Observador	Observador
	Participante	Participante
	Instrumento	Instrumento

Volver al principio

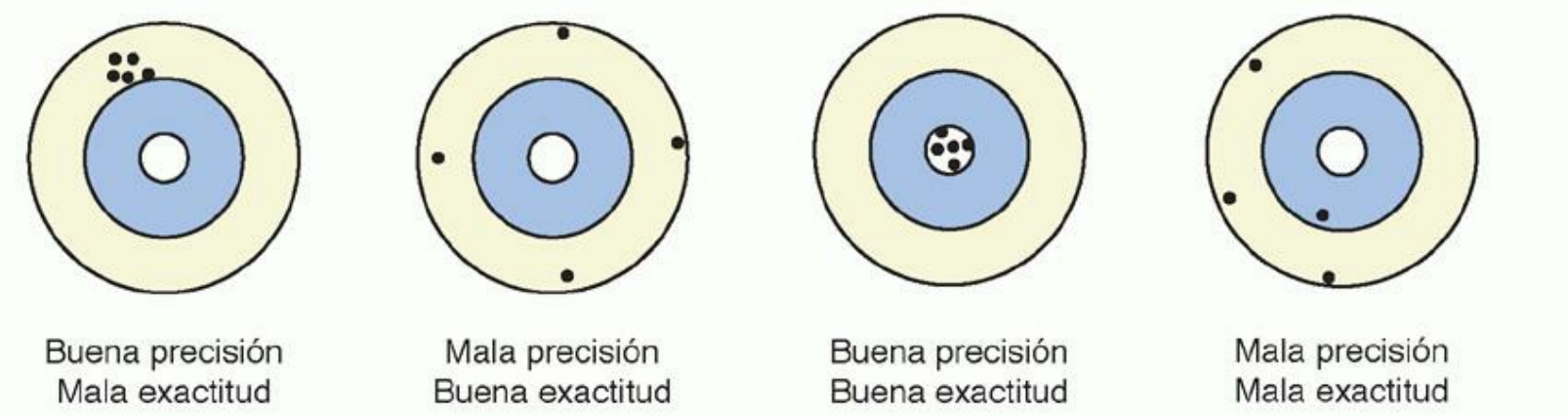
▪ EXACTITUD

La exactitud de una variable es el grado en que representa el valor verdadero.

La exactitud es diferente de la precisión en los aspectos que se muestran en la tabla 4-3 , y las dos no están necesariamente unidas. Si se midiera repetidamente el colesterol sérico usando patrones que se han diluido inadvertidamente dos veces, por ejemplo, el resultado sería inexacto, pero podría seguir siendo preciso (consistentemente por un factor de 2). Este concepto se ilustra también en la figura 4-2 . Sin embargo, exactitud y precisión a menudo van de la mano, en el sentido de que muchas de las estrategias para aumentar la precisión también mejorarán la exactitud.

La exactitud depende del error sistemático (sesgo); cuanto mayor sea el error, menos exacta será la variable. Cada una de las tres clases principales de error de medición observadas en la sección anterior sobre la precisión tiene su equivalente aquí:

- Sesgo del observador. Es una deformación, consciente o inconsciente, de la percepción o la notificación de la medida por el observador. Puede representar errores sistemáticos en el modo en que se utiliza un instrumento, como la tendencia a redondear a la baja las determinaciones de la presión arterial, o utilizar preguntas dirigidas en la entrevista a un participante.
- Sesgo del instrumento. Puede deberse al funcionamiento defectuoso de un instrumento mecánico. Una balanza que no ha sido calibrada recientemente puede haberse desviado a la baja, produciendo continuamente lecturas de pesos corporales bajas.
- Sesgo del participante. Es la deformación de la medición por parte del participante en el estudio, por ejemplo, al notificar un suceso (sesgo de respuesta o de recuerdo). Las pacientes con cáncer de mama que creen que el alcohol es una causa de su cáncer, por ejemplo, pueden exagerar el consumo que refieren.



▪FIGURA 4-2. Diferencia entre precisión y exactitud.
Herramientas de imágenes

La exactitud de una medición se evalúa mejor comparándola, cuando es posible, con un «criterio de referencia»: una medición de referencia realizada por un técnico que se piensa que es la que mejor representa el verdadero valor de la característica. La decisión sobre qué abordaje de la medición se debe considerar como criterio de referencia puede ser un juicio difícil que tiene que hacer el investigador, basándose en el trabajo previo en ese campo.

Para las mediciones en una escala continua, el grado de exactitud se puede expresar como la diferencia media entre la medición en investigación y el criterio de referencia en los distintos participantes en el estudio. Para las mediciones en una escala dicotómica, la exactitud en comparación con el criterio de referencia se puede describir con la sensibilidad y la especificidad (cap. 12). Para las mediciones en escalas categóricas con más de dos opciones de respuesta, se puede calcular el porcentaje de valores correctos con cada método.

Estrategias para mejorar la exactitud

Los principales abordajes para incrementar la exactitud incluyen las primeras cuatro estrategias que se han señalado más arriba para la precisión, y otras tres adicionales (tabla 4-4):

- Normalización de los métodos de medición.
- Formación y certificación de los observadores.
- Perfeccionamiento de los instrumentos.
- Automatización de los instrumentos.
- Realización de medidas que no se perciban. En ocasiones es posible diseñar mediciones de las que no sea consciente el participante, lo que permite eliminar la posibilidad de que conscientemente introduzca sesgo en la variable. Por ejemplo, en una evaluación del efecto de la colocación de un producto para limpiar las manos y un póster sobre la higiene de las manos en la cafetería de un hospital, se utilizaron observadores que se entremezclaban con los clientes de la cafetería (6).
- Calibración del instrumento. La exactitud de muchos instrumentos, especialmente los que son mecánicos o eléctricos, puede aumentar mediante la calibración periódica con un patrón de referencia.
- Enmascaramiento. Esta estrategia clásica no garantiza la exactitud total de las mediciones, aunque puede eliminar el sesgo diferencial que afecta a un grupo del estudio más que a otro. En un estudio clínico con doble enmascaramiento, los pacientes y los observadores desconocen si se ha asignado el fármaco activo o el placebo, y todas las inexactitudes en la medición del criterio de valoración serán las mismas en los dos grupos.

La decisión del interés que se debe poner en aplicar cada una de estas siete estrategias para cada una de las mediciones se basa, como ya se ha señalado en el caso de la precisión, en el juicio del investigador. Las consideraciones son las posibles consecuencias que tendrá el grado previsto de inexactitud sobre las conclusiones del estudio, y la viabilidad y el coste de la estrategia. Las dos primeras estrategias (normalización y formación) se deben utilizar siempre, la calibración es necesaria en cualquier instrumento que pueda cambiar a lo largo del tiempo, y el enmascaramiento es esencial siempre que sea posible.

Volver al principio

▪ VALIDEZ

La validez es similar a la exactitud, aunque nos gusta pensar que añade una dimensión cualitativa a la consideración del grado en que una medición representa el fenómeno de interés. Por ejemplo, las mediciones de la creatinina y la cistatina C en la sangre, dos productos químicos excretados por los riñones, podrían tener la misma *exactitud* (p. ej., a menos del 1% de la concentración verdadera), pero la cistatina C puede ser más *válida* como medición del funcionamiento renal, porque la concentración de creatinina también depende de la cantidad de músculo (7). En la figura 4-2 podemos pensar que la validez describe si el centro de la diana está en la diana correcta.

ESTRATEGIA PARA REDUCIR EL ERROR SISTEMÁTICO	ORIGEN DEL ERROR SISTEMÁTICO	EJEMPLO DE ERROR SISTEMÁTICO	EJEMPLO DE ESTRATEGIA PARA EVITAR EL ERROR

1.	Normalización de los métodos de medición en un manual de instrucciones	Observador	Presión arterial (PA) diastólica elevada de manera constante por el uso del punto en que los ruidos se apagan	Especificar la definición operativa de la PA diastólica como el punto en que los ruidos dejan de oírse
		Participante	Lecturas elevadas de manera constante por medir la PA tras subir las escaleras hasta la consulta	Especificar que el paciente se siente en una habitación tranquila durante 5 min antes de la medición
2.	Formación y aprobación del observador	Observador	PA elevada de manera constante por no seguir procedimientos especificados en el manual de instrucciones	El formador comprueba la exactitud de la lectura del observador con un estetoscopio
3.	Precisión del instrumento	Instrumento	Lecturas de PA elevadas de manera constante con manguitos habituales en pacientes con brazos muy grandes	Usar manguitos de PA de mayor tamaño en pacientes obesos
4.	Automatización del instrumento	Observador	Tendencia consciente o inconsciente del observador a lecturas de PA menores en el grupo aleatorizado al tratamiento activo	Usar dispositivos de medición de PA automáticos
		Participante	Aumento de la PA por la proximidad de un técnico atractivo	Usar dispositivos de medición de PA automáticos
5.	Realización de mediciones discretas	Participante	Tendencia del participante a sobrevalorar el cumplimiento con el fármaco del estudio	Medir la concentración urinaria del fármaco del estudio
6.	Calibración del instrumento	Instrumento	Lecturas de PA consistentemente elevadas por no tener bien calibrado el manómetro	Calibrar cada mes
7.	Enmascaramiento	Observador	Tendencia consciente o inconsciente del observador a leer valores de PA inferiores en el grupo de tratamiento activo	Usar placebo con doble enmascaramiento para ocultar la asignación del grupo de estudio
		Participante	Tendencia del participante que sabía que recibía el fármaco activo a comunicar en exceso los efectos adversos	Usar placebo con doble enmascaramiento para ocultar la asignación del grupo de tratamiento

La validez a menudo no se puede evaluar con un patrón de referencia, particularmente en el caso de mediciones dirigidas a fenómenos subjetivos y abstractos como el dolor y la calidad de vida. Los científicos sociales han creado constructos cualitativos y cuantitativos para abordar la validez de estos abordajes de medición.

- Validez del contenido. Examina hasta qué punto la evaluación representa todos los aspectos de los fenómenos en estudio; por ejemplo, incluyendo preguntas sobre la capacidad funcional social, física, emocional e intelectual para evaluar la calidad de vida.
- Validez aparente. Parece inherentemente razonable, como la medición del dolor en una escala de 10 puntos o la clase social por los ingresos del hogar.
- Validez de constructo. Es el grado en el que un dispositivo de medición específico concuerda con un constructo teórico; por

ejemplo, una prueba de cociente intelectual debe distinguir entre personas que, según la teoría u otras medidas, tienen diferentes niveles de inteligencia.

- Validez predictiva. Es la capacidad que tiene la medición de predecir un resultado; por ejemplo, en qué medida un cuestionario diseñado para evaluar la depresión predice la pérdida de trabajo o el suicidio.

- Validez relacionada con el criterio. Es el grado en que una nueva medida se relaciona con medidas existentes y aceptadas.

El método general para medir fenómenos subjetivos y abstractos es comenzar haciendo una búsqueda en la bibliografía y consultando con expertos en un intento de encontrar un instrumento adecuado (generalmente un cuestionario) que ya haya sido validado. El uso de un instrumento como ese tiene la ventaja de hacer los resultados de un nuevo estudio comparables a trabajos anteriores dentro del mismo campo, y puede simplificar y reforzar el proceso de aplicación para financiación y publicación de los resultados. Sus inconvenientes, no obstante, son que el proceso de validación puede haber sido subóptimo, y que un instrumento retirado de la estantería puede estar anticuado y no ser adecuado para la pregunta de la investigación.

Si los instrumentos de que se dispone no son adecuados para las necesidades del estudio, el investigador puede decidir desarrollar un nuevo método de medición y validarlo por sí mismo. Puede ser un reto interesante e incluso llevar a una contribución notable a la bibliografía, aunque generalmente hace falta mucho tiempo y esfuerzo (cap. 15). Es justo decir que el proceso es, a menudo, menos concluyente de lo que la palabra «validación» implica.

Volver al principio

▪ OTRAS CARACTERÍSTICAS DE LOS MÉTODOS DE MEDICIÓN

Las mediciones deben ser lo suficientemente sensibles como para detectar diferencias en una característica que sea importante para el investigador. El grado de sensibilidad que se precisa depende de la pregunta del estudio. Por ejemplo, un estudio sobre si un nuevo fármaco ayuda a las personas a dejar de fumar podría usar una medición del resultado que no sea muy sensible al *número* de cigarrillos fumados diariamente. Por otro lado, si la pregunta fuera el efecto que la reducción del contenido de nicotina de los cigarrillos tiene sobre el número de cigarrillos fumados, el método sería sensible a diferencias en los hábitos diarios de tan solo algunos cigarrillos.

Una medición ideal es específica, de manera que representa solo las características de interés. La concentración de monóxido de carbono en el aire espirado es una medida del hábito tabáquico que es tan solo moderadamente específica, porque también puede verse afectada por otras exposiciones, como los gases de los automóviles. La especificidad global de la evaluación del hábito tabáquico puede aumentarse añadiendo otras mediciones (como la descripción del propio paciente y la concentración de nicotina sérica) que no están afectadas por la contaminación del aire.

Las mediciones deben ser adecuadas para los objetivos del estudio. Un estudio sobre el estrés como antecedente del infarto de miocardio, por ejemplo, debería tener en cuenta qué tipo de estrés (psicológico o físico, agudo o crónico) tiene interés antes de establecer las definiciones operativas para medirlo.

Las mediciones deben proporcionar una adecuada distribución de las respuestas en la muestra en estudio. Una medida del estado funcional es más útil si produce valores que oscilan desde elevado, en algunos participantes, hasta bajo, en otros. Un motivo importante para la realización de comprobaciones previas es asegurarse de que las respuestas reales no se agrupan todas alrededor de uno de los extremos de un posible intervalo de respuesta (cap. 17).

Siempre que sea posible, las mediciones se deben diseñar de tal manera que se minimicen los juicios subjetivos. La objetividad se consigue reduciendo la intervención del observador y utilizando instrumentos automáticos. Un peligro de estas estrategias, sin embargo, es la consiguiente visión en túnel, que limita el ámbito de las observaciones y la capacidad de descubrir fenómenos no anticipados. Esto se puede corregir incluyendo algunas preguntas abiertas y dando la oportunidad para adquirir datos subjetivos y cualitativos, además de las principales medidas objetivas y cuantitativas.

Cuando se diseña un estudio, existe la tendencia a seguir añadiendo elementos que no son fundamentales para la pregunta de la investigación, pero que *podrían* tener interés. Es cierto que las mediciones adicionales incrementan la probabilidad de obtener hallazgos interesantes, incluyendo algunos que no se habían previsto el comienzo. Sin embargo, es importante tener en mente el valor de la eficiencia y la parsimonia. Se debe diseñar el conjunto completo de mediciones para obtener datos útiles con un coste asequible en cuanto a tiempo y dinero. La obtención de demasiada información es un error frecuente que puede agotar a los participantes, superar al equipo que realiza las mediciones, y dificultar la gestión y el análisis de los datos. La consecuencia puede ser un estudio más costoso, que, paradójicamente, tiene menos éxito a la hora de responder las principales preguntas de la investigación.

Volver al principio

▪ MEDICIONES SOBRE MATERIALES ALMACENADOS

La investigación clínica conlleva realizar mediciones sobre personas en muchos dominios. Algunas de estas mediciones solo pueden realizarse durante un contacto con el participante del estudio, pero muchas pueden llevarse a cabo más tarde sobre muestras biológicas almacenadas para análisis químico y genético, o sobre imágenes de procedimientos radiográficos u otros archivados electrónicamente (tabla 4-5).

Una ventaja de este almacenamiento es la oportunidad de reducir el coste del estudio realizando mediciones solo en personas que, durante el seguimiento de control, tengan un criterio de valoración de interés. Un método magnífico para hacer esto es el diseño de casos y testigos anidado (cap. 8), especialmente si se pueden obtener mediciones pareadas con enmascaramiento en un único lote analítico, lo que elimina el componente de error aleatorio inducido por las diferencias entre lotes. Este abordaje también tiene la ventaja de que los avances científicos varios años después del comienzo del estudio pueden llevar a nuevas ideas y técnicas de medición que se puedan emplear posteriormente, financiadas por subvenciones recién solicitadas.

El creciente interés en la investigación aplicada (cap. 2) aprovecha nuevas medidas que han extendido ampliamente la investigación clínica, por ejemplo, en las áreas de genética y epidemiología molecular (8 , 9) y los estudios de imagen. Las mediciones en muestras que contienen ADN (p. ej., saliva y sangre) pueden proporcionar información sobre genotipos que contribuyen a la aparición de enfermedades o modifican la respuesta de los pacientes al tratamiento. Las determinaciones séricas pueden usarse para estudiar causas moleculares o consecuencias de la enfermedad; por ejemplo, los marcadores inflamatorios ofrecen información útil sobre la fisiopatología de muchas enfermedades. Es importante consultar con expertos sobre los tubos de recogida de muestras y las condiciones de almacenamiento adecuados, con el fin de mantener la calidad de las muestras y hacer que estén disponibles para una amplia variedad de usos posteriores. También es importante obtener el consentimiento informado de los participantes, que incluya la extensión de los posibles usos de las muestras.

TIPO DE MEDICIÓN	EJEMPLOS	MATERIAL PARA MEDICIÓN POSTERIOR
Historia clínica	Diagnósticos, tratamientos, operaciones, síntomas, hallazgos físicos	Historias clínicas en papel o electrónicas
Factores psicosociales	Depresión, antecedentes familiares	Grabaciones de voz, cintas de vídeo
Antropométricas	Altura, peso, composición corporal	Fotografías
Medidas bioquímicas	Colesterol sérico, fibrinógeno plasmático	Suero, plasma, orina, muestras de anatomía patológica
Pruebas genéticas/moleculares	Polimorfismos mononucleotídicos	ADN
Imagen	Densidad ósea, calcio en coronarias	Rayos X, TC, RM
Electromecánicas	Arritmia, cardiopatía congénita	Electrocardiograma, ecocardiograma

Volver al principio

▪ RESUMEN

- Las variables pueden ser numéricas y categóricas. Las variables numéricas pueden ser continuas (se cuantifican en una escala infinita) o discretas (se cuantifican en una escala finita, como los números enteros); las variables categóricas pueden ser nominales (desordenadas) u ordinales (ordenadas), y las que tienen tan solo dos categorías se denominan dicotómicas.

- Las variables que contienen más información confieren más potencia o permiten menores tamaños de la muestra, de acuerdo con la siguiente jerarquía: variables continuas > variables discretas numéricas > variables ordinales > variables nominales y dicotómicas.

- La precisión de una determinación (es decir, la reproducibilidad de medidas repetidas) es otro importante determinante de la potencia y el tamaño de la muestra. Se ve disminuida por el error aleatorio (azar) debido a tres fuentes de variabilidad: del observador, del participante y del instrumento.

- Las estrategias para aumentar la precisión que deben formar parte de cada estudio son definir de forma operativa y normalizar los métodos en un manual operativo. Otras estrategias que con frecuencia son útiles son la formación y la certificación de los observadores, el perfeccionamiento y la automatización de los instrumentos, y la repetición (utilizar la media de mediciones repetidas).

- La exactitud de una medición es el grado en el que se aproxima a un patrón de referencia. La exactitud disminuye por el error sistemático (sesgo) a partir de las tres mismas fuentes: observador, participante e instrumento.

- Las estrategias para aumentar la exactitud incluyen todas las enumeradas con respecto a la precisión, con la excepción de la repetición. Además, la exactitud se fomenta por medidas no molestas, calibración y (en comparaciones entre grupos) enmascaramiento.

- La validez es el grado en el que una medida representa los fenómenos de pretende medir; se utiliza habitualmente para variables más abstractas y subjetivas, y se evalúan mediante la validez de contenido, la validez aparente, la validez de constructo, la validez predictiva y la validez relacionada con el criterio.

- Las mediciones individuales deben ser sensibles, específicas, apropiadas y objetivas, y deben dar lugar a un intervalo de valores. En conjunto, deben ser amplias pero parcas, atendiendo a la pregunta de la investigación con un coste moderado en cuanto a tiempo y dinero.

- Los investigadores deben considerar almacenar imágenes y otros materiales para posteriores determinaciones que puedan aprovecharse de nuevas tecnologías a medida que se desarrollen, y de la eficacia de diseños de casos y testigos anidados.

Volver al principio

[1]Cuando hay dos mediciones de una variable continua por participante, puede ser tentador expresar su concordancia utilizando un coeficiente de correlación. Sin embargo, como el coeficiente de correlación es muy sensible a los valores extremos (3 , 4), un abordaje mejor es el «gráfico de Bland y Altman», en el que se representa la diferencia entre las dos mediciones en función de su media. Si el valor absoluto de la diferencia entre las mediciones tiende a aumentar linealmente con la media, el coeficiente de correlación es una forma mejor de medir la variabilidad que la desviación típica intrapaciente.

APÉNDICE 4
NA

▪ DEFINICIÓN OPERATIVA DE UNA MEDICIÓN DE LA FUERZA DE PRENSIÓN

El manual operativo describe el método para medir y registrar los resultados de todas las determinaciones que se realizan en el estudio. Este ejemplo, el del manual operativo del *Study of Osteoporotic Fractures* (Estudio de fracturas osteoporóticas), describe el uso de un dinamómetro para medir fuerza de prensión. Para normalizar las instrucciones de examinador a examinador y de participante a participante, el protocolo incluye un documento (guion) de instrucciones para que el participante lea palabra por palabra.

▪ PROTOCOLO PARA MEDIR FUERZA DE PRENSIÓN CON EL DINAMÓMETRO

La fuerza de prensión debe medirse en las dos manos. El mango debe ajustarse de modo que el participante sostenga el dinamómetro

confortablemente. Coloque el dinamómetro en la mano derecha con el dial dirigido hacia la palma. El brazo del participante debe estar flexionado 90° en el codo, con el antebrazo paralelo al suelo.

- Demostrar la prueba al participante. Mientras se demuestra, use la siguiente descripción: «Este aparato mide la fuerza del brazo y la parte superior del cuerpo. Mediremos la fuerza de prensión en ambos brazos. Le demostraré cómo se realiza. Doble el codo formando un ángulo de 90°, con el antebrazo paralelo al suelo. No deje que el brazo toque el costado. Baje el aparato y apriete lo más fuerte que pueda mientras cuento hasta tres. Una vez que el brazo esté totalmente extendido, puede aflojar su prensión».

- Permita un ensayo con cada brazo, empezando con el derecho si el participante es diestro. En el segundo ensayo, registre los kilogramos de fuerza del dial hasta el punto de 0,5 kg más próximo.

- Vuelva a colocar en cero el dial. Repita el procedimiento con el otro brazo.

El brazo no debe entrar en contacto con el cuerpo. La acción de prensión debe ser un apretón lento y sostenido, en lugar de un apretón explosivo.

BIBLIOGRAFÍA

1. Michaelsson K, Baron JA, Snellman G, et al. Plasma vitamin D and mortality in older men: a community-based prospective cohort study. Am J Clin Nutr 2010;92:841-848. Citado aquí

2. Ware JE, Gandek B Jr. Overview of the SF-36 health survey and the International Quality of Life Assessment Project. J Clin Epidemiol 1998;51:903-912. Citado aquí

3. Bland JM, Altman DG. Measurement error and correlation coefficients. BMJ 1996;313:41-42; also, Measurement error proportional to the mean. BMJ 1996;313:106. Citado aquí

4. Newman TB, Kohn M. Evidence-based diagnosis. New York: Cambridge University Press, 2009. Citado aquí

5. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37- 46. Citado aquí

6. Filion K, Kukanich KS, Chapman B, et al. Observation-based evaluation of hand hygiene practices and the effects of an intervention at a public hospital cafeteria. Am J Infect Control 2011;39:464-470. Citado aquí

7. Peralta CA, Shlipak MG, Judd S, et al. Detection of chronic kidney disease with creatinine, cystatin C, and urine albumin-to-creatinine ratio and association with progression to end-stage renal disease and mortality. JAMA 2011;305:1545-1552. Citado aquí

8. Guttmacher AE, Collins FS. Genomic medicine: a primer. NEJM 2002;347:1512-1520. Citado aquí

9. Healy DG. Case-control studies in the genomic era: a clinician's guide. The Lancet Neurology 2006;5:701-707. Citado aquí

CAPÍTULO 5 Preparación para el cálculo del tamaño de la muestra: hipótesis y principios subyacentes

Warren S. Browner
Thomas B. Newman
Stephen B. Hulley

Después de que el investigador ha decidido a quién y qué va a estudiar, y el diseño que va a usar, deberá decidir cuántos participantes integrarán la muestra. Incluso el estudio de ejecución más rigurosa puede fracasar en la respuesta a su pregunta de investigación si el tamaño de la muestra es demasiado pequeño. Por otro lado, un estudio con una muestra demasiado grande resultará más difícil y costoso de lo necesario. El objetivo de la planificación del tamaño de la muestra consiste en calcular un número adecuado de participantes para un diseño de estudio concreto.

Aunque constituyen una guía útil, los cálculos del tamaño de la muestra dan una impresión engañosa de objetividad estadística. Son solo igual de exactos que los datos y cálculos en los que se basan, que, a menudo, son simplemente supuestos informados. Debe considerarse que la planificación del tamaño de la muestra es una forma matemática de realizar un cálculo aproximado. A menudo demuestra que el diseño de investigación no es posible o que se necesitan diferentes variables predictivas o de respuesta. Por lo tanto, el tamaño de la muestra debe calcularse al principio de la fase de diseño de un estudio, cuando todavía es posible realizar cambios importantes.

Antes de establecer los métodos específicos para calcular el tamaño de la muestra para varios diseños de investigación habituales, en el capítulo 6, dedicaremos algún tiempo a considerar los principios subyacentes. Los lectores que encuentren algunos de estos principios confusos disfrutarán descubriendo que la planificación del tamaño de la muestra no requiere un dominio total. Sin embargo, igual que una receta tiene más sentido si el cocinero está algo familiarizado con los ingredientes, los cálculos del tamaño de la muestra son más senillos si el investigador conoce los conceptos básicos. Incluso si tiene previsto pedir a un amigo bioestadístico que calcule el tamaño de la muestra de su estudio, tener algunos conocimientos de cómo funciona el proceso le permitirá participar más activamente en la consideración de los supuestos y estimaciones implicados en el cálculo.

▪ HIPÓTESIS

El proceso comienza reformulando la pregunta de la investigación como hipótesis de investigación, que resume los principales elementos del estudio: la muestra y las variables predictivas y de respuesta. Por ejemplo, suponga que su pregunta de la investigación es que las personas que realizan crucigramas tienen menos probabilidad de presentar demencia. Su hipótesis de investigación tendría que especificar la muestra (p. ej., personas que viven en una comunidad de jubilados que tienen una función cognitiva normal), la variable predictiva (realizar crucigramas al menos una vez a la semana en promedio) y la variable de respuesta (una puntuación anómala en una prueba estandarizada de función cognitiva después de 2 años de seguimiento).

Las hipótesis en sí mismas no son necesarias en los estudios descriptivos, que describen cómo se distribuyen las características en una población, como la prevalencia de una función cognitiva anómala en la comunidad de jubilados. (Esto no significa, sin embargo, que no vaya a necesitar calcular el tamaño de una muestra para un estudio descriptivo, sino que los métodos para hacerlo, descritos en el capítulo 6, son diferentes.) Se necesitan las hipótesis en estudios que usarán pruebas de significación estadística para comparar hallazgos entre grupos, como si los ancianos que realizan crucigramas con frecuencia tienen menos probabilidad de llegar a tener demencia. Debido a que la mayor parte de los estudios de observación y todos los estudios experimentales plantean preguntas de investigación que conllevan la realización de comparaciones, la mayor parte de los estudios necesita especificar, al menos, una hipótesis. Si alguno de los siguientes términos aparece en la pregunta de la investigación, el estudio no será simplemente descriptivo, y deberá formularse una hipótesis de la investigación: mayor que, menor que, más probable que, asociado a, comparado con, relacionado con, similar, correlacionado con, causa y produce.

Características de una buena hipótesis

Una buena hipótesis debe basarse en una buena pregunta de investigación. También debe ser sencilla y específica, y se debe describir por adelantado.

Sencilla frente a compleja

Una hipótesis sencilla contiene una variable predictiva y una de respuesta (o dependiente):

En pacientes con diabetes de tipo II, un estilo de vida sedentario se asocia a mayor riesgo de presentar proteinuria.

Una hipótesis compleja contiene más de una variable predictiva:

En pacientes con diabetes de tipo II, un estilo de vida sedentario y el consumo de alcohol se asocian a mayor riesgo de presentar proteinuria.

O más de una variable de respuesta:

En pacientes con diabetes de tipo II, el consumo de alcohol se asocia a mayor riesgo de presentar proteinuria y neuropatía.

Las hipótesis complejas como estas no se comprueban fácilmente con una sola prueba estadística, y se abordan más fácilmente como dos o más hipótesis sencillas. A veces, no obstante, puede utilizarse una variable predictiva o de respuesta combinada:

En pacientes con diabetes de tipo II, el consumo de alcohol se asocia a mayor riesgo de presentar una complicación microvascular (como proteinuria, nefropatía o retinopatía).

En este ejemplo, el investigador ha decidido qué ocurre si un participante tiene una complicación, no qué tipo de complicación se produce.

Específica frente a imprecisa

Una hipótesis específica no deja ambigüedad sobre los participantes y las variables, ni sobre cómo se aplicará la prueba de significación estadística. Usa definiciones operativas concisas, que resumen la naturaleza y el origen de los participantes y el modo en que se medirán las variables:

El uso previo de antidepresivos tricíclicos durante al menos 6 semanas es más frecuente en pacientes ingresados por infarto de miocardio en el hospital Longview que en los testigos ingresados por neumonía.

La frase es larga, pero comunica la naturaleza del estudio de una forma clara, que reduce al mínimo cualquier oportunidad de probar algo que sea un poco diferente una vez que se han examinado los hallazgos del estudio. Sería incorrecto sustituir, durante la fase de análisis del estudio, una medida diferente de la variable predictiva, como la depresión comunicada por el propio paciente, sin considerar el tema del estudio de múltiples hipótesis (un punto que se comentará al final del capítulo). Generalmente, para mantener concisa la hipótesis de la investigación, algunos de estos detalles se dejan claros en el plan de estudio, en lugar de establecerse en la hipótesis de investigación. Pero deben estar siempre claros en la concepción del estudio por parte del investigador, y deben explicarse en el protocolo.

Con frecuencia, es evidente, a partir de la hipótesis de investigación, si la variable predictiva y la variable de respuesta son dicotómicas, continuas o categóricas. Si no está claro, puede especificarse el tipo de variables:

En hombres no obesos de 35 a 59 años de edad, la participación al menos una vez a la semana en una liga de bolos se asocia a mayor riesgo de presentar obesidad (índice de masa corporal > 30 kg/m^2) durante 10 años de seguimiento.

Una vez más, si la hipótesis de investigación llega a ser difícil de manejar, pueden omitirse las definiciones siempre que se aclaren en algún otro lugar.

Anticipadamente frente a con posterioridad

La hipótesis debe plantearse por escrito al principio del estudio. Esto mantendrá el esfuerzo de la investigación centrado en el objetivo primario. Una sola hipótesis preestablecida también crea una base más fuerte para interpretar los resultados del estudio que varias hipótesis que surgen como resultado de inspeccionar los datos. Las hipótesis que se formulan tras el examen de los datos son una forma de probar múltiples hipótesis, lo que puede llevar a asignar una importancia excesiva a los hallazgos.

Hipótesis nula y alternativa

Advertencia: si no ha recibido nunca educación formal en estadística, o si ha olvidado lo que aprendió, los próximos párrafos pueden no tener sentido las primeras veces que los lea. Intente superar la terminología, aunque parezca difícil o estúpida.

El proceso comienza reformulando la hipótesis de la investigación para tener una hipótesis que proponga que no hay diferencias entre los grupos que se comparan. Esta reformulación, denominada hipótesis nula, será la base formal del estudio de la significación estadística cuando analice los datos al final del estudio. Al aceptar que realmente no hay asociación en la población, las pruebas estadísticas ayudarán a estimar la probabilidad de que una asociación observada en un estudio se pueda deber al azar.

Por ejemplo, suponga que su pregunta de la investigación es si beber agua del grifo no purificada se asocia a mayor riesgo de presentar enfermedad ulcerosa péptica (tal vez debido a una mayor probabilidad de contaminación por *H. pylori*). La hipótesis nula, que no hay asociación entre las variables predictiva y de respuesta en la población, sería:

Las personas de Phnom Penh que beben agua del grifo tienen el mismo riesgo de presentar enfermedad ulcerosa péptica que las que beben agua embotellada.

La proposición de que hay una asociación («Las personas de Phnom Penh que beben agua del grifo tienen mayor riesgo de presentar enfermedad ulcerosa péptica que las que beben agua embotellada») se denomina hipótesis alternativa. La hipótesis alternativa no se puede estudiar directamente; se acepta por defecto si la prueba de significación estadística rechaza la hipótesis nula (v. más adelante).

Hacen falta otros términos confusos. La hipótesis alternativa puede ser unilateral o bilateral. Una hipótesis alternativa unilateral especifica la dirección de la asociación entre las variables predictiva y de respuesta. La hipótesis de que beber agua del grifo aumenta el riesgo de enfermedad ulcerosa péptica (en comparación con el agua embotellada) es una hipótesis unilateral. Una hipótesis alternativa bilateral afirma únicamente que hay una asociación; no especifica la dirección. Por ejemplo, «beber agua del grifo se asocia a un riesgo diferente de enfermedad ulcerosa péptica (mayor o menor) que beber agua embotellada».

Las hipótesis unilaterales pueden ser apropiadas en determinadas circunstancias, como cuando solo es clínicamente importante o biológicamente significativa una dirección para una asociación. Un ejemplo lo constituye la hipótesis unilateral de que un nuevo fármaco para la hipertensión tiene más probabilidad de causar exantema que un placebo; no suele merecer la pena comprobar la posibilidad de que el fármaco cause menos exantema que el placebo (sin embargo, lo merecería si el fármaco tuviera propiedades antiinflamatorias). Una hipótesis unilateral también puede ser apropiada cuando existen pruebas importantes, a partir de estudios previos, de que es improbable que se produzca una asociación en una de las dos direcciones, como en un estudio en el que se comprobó si fumar cigarrillos afecta al riesgo de cáncer cerebral. Debido a que el tabaquismo se ha asociado a mayor riesgo de muchos tipos diferentes de cáncer, podría bastar una hipótesis alternativa unilateral (p. ej., que fumar aumenta el riesgo de cáncer cerebral). Sin embargo, los investigadores deben conocer que muchas hipótesis bien fundamentadas (p. ej., que el tratamiento con β caroteno disminuirá el riesgo de cáncer de pulmón, o que el tratamiento con fármacos que reducen el número de extrasístoles ventriculares disminuirá la muerte súbita entre los pacientes con arritmias ventriculares) se tornarán erróneas cuando se estudien en ensayos aleatorizados. Efectivamente, en estos dos ejemplos, los resultados de ensayos bien realizados revelaron un efecto estadísticamente significativo de dirección opuesta a la que los investigadores esperaban encontrar (1 , 2 , 3). En general, creemos que la mayoría de las hipótesis alternativas deben ser bilaterales.

Es importante tener en cuenta la diferencia entre la hipótesis de investigación, que habitualmente es unilateral, y la hipótesis alternativa, que se usa cuando se planifica el tamaño de la muestra y que casi siempre es bilateral. Por ejemplo, suponga que la hipótesis de investigación es que el uso recurrente de antibióticos durante la infancia se asocia a mayor riesgo de enfermedad inflamatoria intestinal. Esa hipótesis especifica la dirección del efecto anticipado, por lo que es unilateral. ¿Por qué usar una hipótesis alternativa bilateral al planificar el tamaño de la muestra? La respuesta está en que la mayor parte de las veces ambos lados de la hipótesis alternativa (es decir, mayor riesgo o menor riesgo) son interesantes, y los investigadores podrían desear publicar los resultados independientemente de cuál fuera la dirección observada. El rigor estadístico precisa que el investigador escoja entre hipótesis unilaterales y bilaterales antes de analizar los datos; el cambio de una hipótesis bilateral a otra unilateral para reducir el valor de *p* (v. más adelante) no es correcto. Además (y esta es probablemente la razón por la que las hipótesis alternativas bilaterales son mucho más frecuentes), la mayor parte de los revisores de financiaciones y manuscritos esperan hipótesis bilaterales y critican los enfoques unilaterales.

Volver al principio

▪ PRINCIPIOS ESTADÍSTICOS SUBYACENTES

Una hipótesis de investigación, como la que afirma que 15 min o más de ejercicio al día se asocian a menor glucemia media en ayunas en mujeres de mediana edad con diabetes, es verdadera o falsa en el mundo real. Como un investigador no puede estudiar a todas las mujeres de mediana edad con diabetes, deberá comprobar la hipótesis en una muestra de esa población objetivo. Como se señala en la figura 1-5, siempre se necesitará extraer inferencias sobre los fenómenos en la población a partir de sucesos observados en la muestra. Lamentablemente, tan solo por azar, en ocasiones, lo que ocurre en una muestra no refleja lo que habría ocurrido si se hubiera estudiado la población completa.

De algún modo, el problema del investigador es similar al problema al que se enfrenta un jurado que juzga a un acusado (tabla 5-1). La verdad absoluta sobre si el acusado cometió el delito no suele poder determinarse. Antes bien, el jurado empieza presuponiendo la inocencia: el acusado no cometió el delito. El jurado debe decidir si existen pruebas suficientes para rechazar la presunta inocencia del acusado; lo habitual se conoce como más allá de una duda razonable. Un jurado puede, no obstante, equivocarse, condenando a un acusado inocente o no condenando a uno que sea culpable.

De modo similar, el investigador empieza suponiendo la hipótesis nula de ausencia de asociación entre las variables predictiva y de respuesta en la población. Según los datos recopilados en su muestra, usa pruebas estadísticas para determinar si existen pruebas suficientes para rechazar la hipótesis nula a favor de la hipótesis alternativa de que existe una asociación en la población. Lo habitual (la referencia) para estas pruebas se denomina nivel de significación estadística.

Errores de tipo I y de tipo II

Como un jurado, un investigador puede llegar a una conclusión errónea. A veces, por casualidad, tan solo una muestra no es representativa de la población, y los resultados de la muestra no reflejan la realidad en la población, dando lugar a una inferencia errónea. Se produce un error de tipo I (positivo falso) si un investigador rechaza una hipótesis nula que es realmente cierta en la población; se produce un error de tipo II (negativo falso) si el investigador no rechaza una hipótesis nula que realmente es falsa en la población. Aunque estos errores nunca pueden evitarse totalmente, el investigador puede reducir su probabilidad aumentando el tamaño de la muestra (cuanto mayor sea la muestra, menos probable será que difiera notablemente de la población), o manipulando el diseño o las determinaciones de otros modos que se comentarán.

DECISIÓN DE UN JURADO	**PRUEBA ESTADÍSTICA**
Inocencia: el acusado no falsificó dinero	**Hipótesis nula:** no existe asociación entre el caroteno de la dieta y la incidencia de cáncer de colon en la población
Culpabilidad: el acusado falsificó dinero	**Hipótesis alternativa:** existe una asociación entre el caroteno de la dieta y la incidencia de cáncer de colon
Referencia para rechazar la inocencia: más allá de una duda razonable	**Referencia para rechazar la hipótesis nula:** nivel de significación estadística *(a)*
Juicio correcto: condenar a un falsificador	**Inferencia correcta:** concluir que existe una asociación entre el caroteno de la dieta y el cáncer de colon cuando existe en la población
Juicio correcto: absolver a una persona inocente	**Inferencia correcta:** concluir que no existe asociación entre el caroteno y el cáncer de colon cuando no existe
Juicio incorrecto: condenar a una persona inocente	**Inferencia incorrecta (error de tipo I):** concluir que existe una asociación entre el caroteno de la dieta y el cáncer de colon cuando en realidad no hay ninguna
Juicio incorrecto: absolver a un falsificador	**Inferencia incorrecta (error de tipo II):** concluir que no existe asociación entre el caroteno de la dieta y el cáncer de colon cuando en realidad existe

En este capítulo y en el siguiente solo se abordarán modos de reducir los errores de tipo I y de tipo II debidos a la variación por el azar, también conocidos como error aleatorio. También pueden producirse resultados positivos falsos y negativos falsos debido a sesgo, pero estos errores por sesgo no suelen denominarse errores de tipo I y de tipo II. Son errores molestos porque pueden ser difíciles de detectar, y generalmente no pueden cuantificarse usando métodos estadísticos ni evitarse aumentando el tamaño de la muestra. (En los caps. 1, 3, 4 y 7-12, se exponen modos de reducir los errores debidos al sesgo.)

Magnitud del efecto

La probabilidad de que un estudio vaya a ser capaz de detectar una asociación entre una variable predictiva y una variable de respuesta en una muestra depende de la magnitud real de esa asociación en la población. Si es grande (p. ej., una diferencia de 20 mg/dl de la glucemia basal), será fácil detectarlo en la muestra. Por el contrario, si el tamaño de la asociación es pequeño (una diferencia de 2 mg/dl), será difícil detectarlo en la muestra.

Desgraciadamente, el investigador casi nunca conoce la magnitud exacta de la asociación; ¡uno de los objetivos del estudio es calcularla! Más bien, el investigador debe escoger la magnitud de la asociación en la población que desea detectar en la muestra. Esa cantidad es lo que se conoce como magnitud del efecto. Seleccionar una adecuada magnitud del efecto es el aspecto más difícil de la planificación del tamaño de la muestra (4). El investigador debe intentar encontrar datos de estudios anteriores en áreas relacionadas para crear un supuesto informado sobre una razonable magnitud del efecto. Por otro lado, puede escoger el tamaño de efecto más pequeño que, en su

opinión, sería clínicamente significativo (p. ej., una reducción de 10 mg/dl en la glucemia en ayunas).

Por supuesto, desde el punto de vista de la salud pública, incluso una reducción de la glucemia en ayunas de 2 o 3 mg/dl sería importante, especialmente si fuese fácil de conseguir. La elección de la magnitud del efecto es siempre arbitraria, y las consideraciones de viabilidad son siempre de capital importancia. Efectivamente, cuando el número de participantes disponibles o asequibles es bajo, el investigador puede tener que trabajar de forma retrospectiva (cap. 6) para determinar la magnitud del efecto que podrá detectar, dado el número de pacientes que puede estudiar.

Muchos estudios tienen varias magnitudes del efecto, porque miden varias variables predictivas y de respuesta diferentes. Cuando se diseña un estudio, se debe determinar el tamaño de la muestra utilizando la magnitud del efecto deseada para la hipótesis más importante; podrán calcularse entonces las magnitudes del efecto detectables para otras hipótesis. Si existen varias hipótesis de importancia similar, el tamaño de la muestra para el estudio deberá basarse en cualquiera de las hipótesis que necesite la muestra mayor.

α, β y potencia

Tras completar un estudio, el investigador usa pruebas estadísticas para intentar rechazar la hipótesis nula en favor de su alternativa, casi del mismo modo que un fiscal intenta convencer al jurado para que rechace la inocencia a favor de la culpabilidad. Dependiendo de si la hipótesis nula es verdadera o falsa en la población en estudio, y suponiendo que el estudio carezca de sesgo, se pueden producir cuatro situaciones (tabla 5-2). En dos de ellas, los hallazgos en la muestra y en la realidad en la población coinciden, y la inferencia del investigador será correcta. En las otras dos situaciones, se ha producido un error de tipo I o de tipo II, y la inferencia será incorrecta.

El investigador establece, antes de hacer el estudio, la máxima probabilidad que tolerará de realizar errores de tipo I y II. La máxima probabilidad de cometer un error de tipo I (rechazar la hipótesis nula cuando en realidad es cierta) se denomina α (alfa). Otro nombre para α es nivel de significación estadística.

Si, por ejemplo, a un estudio de los efectos del ejercicio sobre la glucemia en ayunas se le asigna un valor α de 0,05, el investigador ha establecido el 5% como la probabilidad máxima de rechazo incorrecto de la hipótesis nula si esta es cierta (con lo que se infiere que el ejercicio y la glucemia en ayunas están asociados en la población cuando, de hecho, no lo están). Este es el nivel de duda razonable que el investigador estará dispuesto a aceptar cuando use pruebas estadísticas para analizar los datos una vez completado el estudio.

La probabilidad de cometer un error de tipo II (no rechazar la hipótesis nula cuando en realidad es falsa) se denomina β (beta). La cantidad (1 - β) se denomina potencia, y es la probabilidad de rechazar correctamente la hipótesis nula en la muestra si el efecto real en la población es igual (o mayor) que la magnitud del efecto.

Si se establece el valor de β en 0,10, el investigador ha decidido que está dispuesto a aceptar un riesgo del 1 % de perder una asociación de una magnitud del efecto concreta si existe. Esto representa una potencia de 0,90, es decir, una posibilidad del 9 % de encontrar una asociación de ese tamaño o mayor. Por ejemplo, suponga que el ejercicio realmente conduce a una reducción promedio de 20 mg/dl de la glucemia en ayunas en mujeres diabéticas de la población. Si el investigador repitiera el estudio con la misma potencia del 9 % en numerosas ocasiones, esperaríamos que en 9 de cada 10 estudios rechazara correctamente la hipótesis nula con el nivel de alfa especificado (0,05), y concluiría que el ejercicio se asocia a la glucemia basal. Esto no significa que el investigador no pudiera detectar una magnitud del efecto menor en la población, por ejemplo, una reducción de 15 mg/dl; sencillamente significa que tendrá una probabilidad menor del 9 % de hacerlo.

De forma ideal, α y β deberían ser próximos a 0, lo que minimizaría la posibilidad de obtener resultados positivos falsos y negativos falsos. Sin embargo, reducirlos requiere que se aumente el tamaño de la muestra, o una de las otras estrategias que se discuten en el capítulo 6. La planificación del tamaño de la muestra pretende escoger una cantidad suficiente de participantes para mantener α y β en un nivel aceptablemente bajo sin que el estudio sea innecesariamente caro y difícil.

	REALIDAD EN LA POBLACIÓN	
RESULTADOS EN LA MUESTRA DEL ESTUDIO	**ASOCIACIÓN ENTRE PREDICTIVA Y DE RESPUESTA**	**SIN ASOCIACIÓN ENTRE PREDICTIVA Y DE RESPUESTA**
Rechazar la hipótesis nula	Correcto	Error de tipo I

No rechazar la hipótesis nula	Error de tipo II	Correcto

Muchos estudios establecen α en 0,05 y β en 0,20 (una potencia de 0,80). Son valores arbitrarios, y a veces se utilizan otros: los límites convencionales para α están entre 0,01 y 0,10, y los de β entre 0,05 y 0,20. En general, el investigador debe usar un valor de α bajo cuando la pregunta de la investigación hace que sea particularmente importante evitar un error de tipo I (positivo falso): por ejemplo, al probar la eficacia de un fármaco que puede ser peligroso. Debe usar un valor de β bajo (y una magnitud del efecto pequeña) cuando es especialmente importante evitar un error de tipo II (negativo falso): por ejemplo, al tranquilizar a las personas de que vivir cerca de un vertedero de residuos tóxicos no es peligroso.

Valor de probabilidad (valor de *p*)

Ahora es el momento de volver a la hipótesis nula, cuya finalidad subyacente finalmente quedará clara. La hipótesis nula tiene una única función: actuar como «hombre de paja». Se supone que es verdadera para poder rechazarla como falsa con una prueba estadística. Cuando se analizan los datos, se utiliza una prueba estadística para determinar el valor de***p***, que es la probabilidad de observar (únicamente por el azar) un efecto tan grande, o mayor, como el que se vería en el estudio si la hipótesis nula realmente fuera cierta. Lo fundamental es reconocer que si la hipótesis nula es verdadera, y si realmente no hay diferencias en la población, entonces la única forma en la que el estudio habría encontrado una diferencia en la muestra sería por azar.

Si esa probabilidad es pequeña, entonces se puede rechazar la hipótesis nula de ausencia de diferencia a favor de la hipótesis alternativa, que sí hay diferencia. Con «pequeño» nos referimos a un valor de *p* menor que a, el nivel de significación estadística predeterminado.

Sin embargo, un resultado «no significativo» (uno con un valor de *p* mayor que α no indica que no existe asociación en la población; solo significa que el resultado observado en la muestra es pequeño comparado con el que podría haberse producido solo por azar. Por ejemplo, un investigador podría haber encontrado que las mujeres que practicaban deportes colegiales tenían el doble de probabilidad de precisar una sustitución total de cadera en fases posteriores de la vida que aquellas que no lo hacían, aunque, como el número de sustituciones de cadera en el estudio fue bajo, este efecto aparente tenía un valor de *p* de tan solo 0,08. Esto significa que incluso si la actividad física y la lesión de cadera no estaban asociadas en la población, habría una probabilidad del 8% de encontrar una asociación al menos tan grande como la que observó el investigador *únicamente por azar*. Si el investigador había establecido el nivel de significación como un valor de α bilateral de 0,05, debería haber concluido que la asociación en la muestra «no era estadísticamente significativa».

Podría ser tentador que el investigador cambiara su idea y pasara a un valor de *p* unilateral, y describiera que «$p = 0,04$». Una posibilidad mejor sería describir los resultados con el intervalo de confianza al 9 %, y comentar que «los resultados, aunque sugestivos de una asociación, no alcanzaron la significación estadística ($p = 0,08$)». Esta solución mantiene la integridad del diseño de hipótesis bilateral original, y también reconoce que la significación estadística no es una situación de todo o nada.

Lados de la hipótesis alternativa

Recuerde que una hipótesis alternativa tiene realmente dos lados, pudiéndose comprobar cada uno o ambos en la muestra usando pruebas estadísticas unilaterales o bilaterales [1] . Cuando se utiliza una prueba estadística bilateral, el valor de *p* incluye la probabilidad de cometer un error de tipo I en cada una de las dos direcciones, lo que supone aproximadamente duplicar la probabilidad en una u otra dirección solamente. Es fácil pasar de un valor de *p* unilateral a un valor de *p* bilateral, y viceversa. Un valor de *p* unilateral de 0,05, por ejemplo, suele ser lo mismo que un valor *p* bilateral de 0,10. (Algunas pruebas estadísticas son simétricas, y es por lo que decimos «suele».)

En esas raras situaciones en las que un investigador solo está interesado en uno de los lados de la hipótesis alternativa (p. ej., un estudio de ausencia de inferioridad diseñado para determinar si un nuevo antibiótico no es menos eficaz que otro que se utiliza actualmente; v. cap. 11), el tamaño de la muestra debe calcularse según ello. Sin embargo, nunca debe usarse una hipótesis unilateral tan solo para reducir el tamaño de la muestra.

Tipo de prueba estadística

Las fórmulas usadas para calcular el tamaño de la muestra se basan en supuestos matemáticos, lo que difiere para cada prueba estadística. Antes de que pueda calcularse el tamaño de la muestra, el investigador debe decidir el enfoque estadístico para analizar los datos. Esa elección depende fundamentalmente del tipo de variables predictivas y de respuesta del estudio. En la tabla 6-1 se enumeran algunos estadísticos frecuentes usados en el análisis de datos, y en el capítulo 6 se ofrecen enfoques simplificados para calcular el tamaño de la muestra para estudios que utilizan estos estadísticos.

Volver al principio

▪ OTROS PUNTOS

Variabilidad

No es simplemente la magnitud de un efecto lo que es importante; su variabilidad también lo es. Las pruebas estadísticas dependen de si son capaces de mostrar una diferencia entre los grupos que se comparan. Cuanto mayor es la variabilidad (o dispersión) de la variable de respuesta entre los participantes, más probable será que se superpongan los valores de los grupos, y más difícil será demostrar una diferencia general entre ellos. Debido a que el error de medida contribuye a la variabilidad general, las medidas menos precisas requieren tamaños de muestra mayores (5).

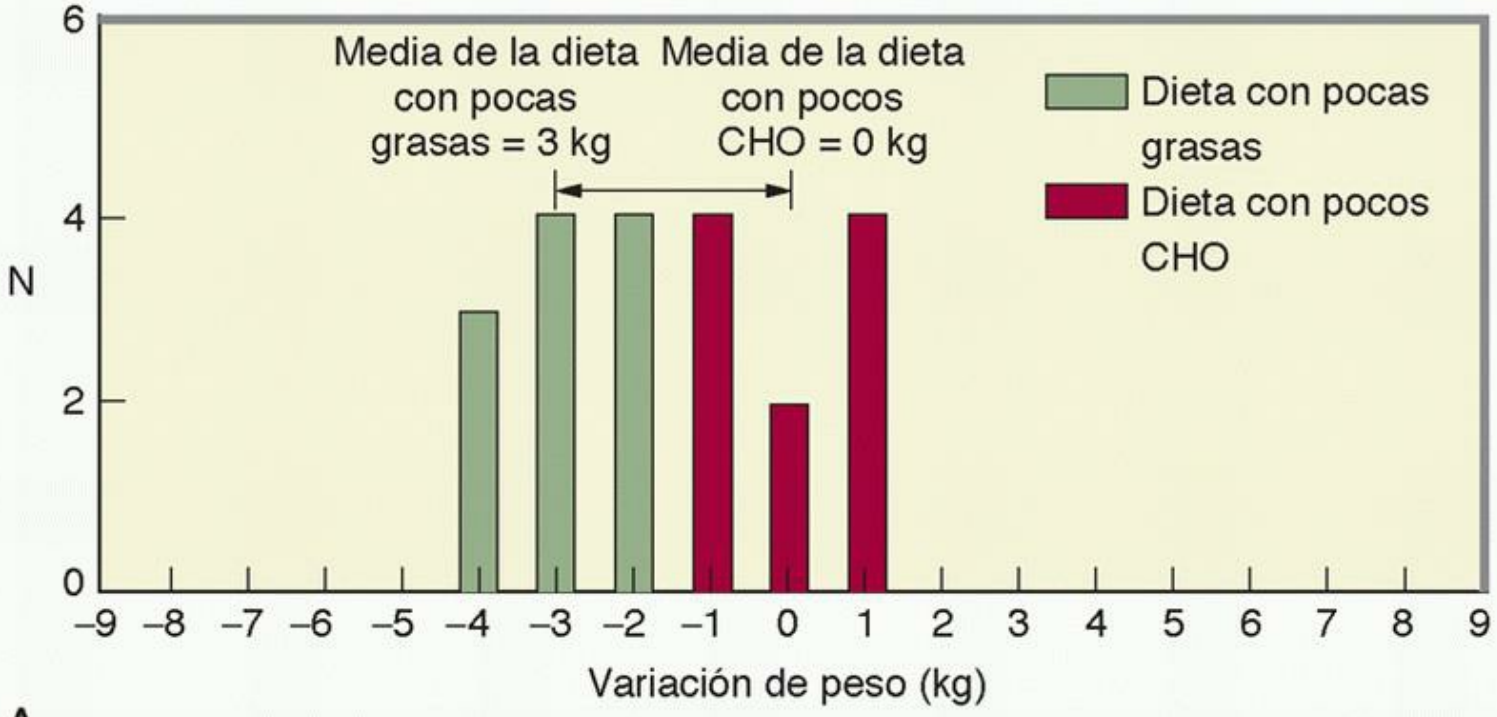
Considere un estudio sobre los efectos de dos dietas (pocas grasas y pocos hidratos de carbono) en la consecución de una pérdida de peso en 20 pacientes obesos. Si todos los que siguen la dieta baja en grasas pierden unos 3 kg y todos los que siguen la dieta con pocos hidratos de carbono pierden poco peso o no pierden peso (una magnitud del efecto de 3 kg), es probable que la primera dieta sea realmente mejor (fig. 5-1A). Por otro lado, si la pérdida promedio de peso es de 3 kg en el grupo con la dieta baja en grasas y de 0 kg en el grupo de la dieta con hidratos de carbono, pero hay mucha superposición entre los dos grupos (la situación de la figura 5-1B), la mayor variabilidad haría que fuera más difícil detectar una diferencia entre las dietas, y haría falta un mayor tamaño de la muestra.

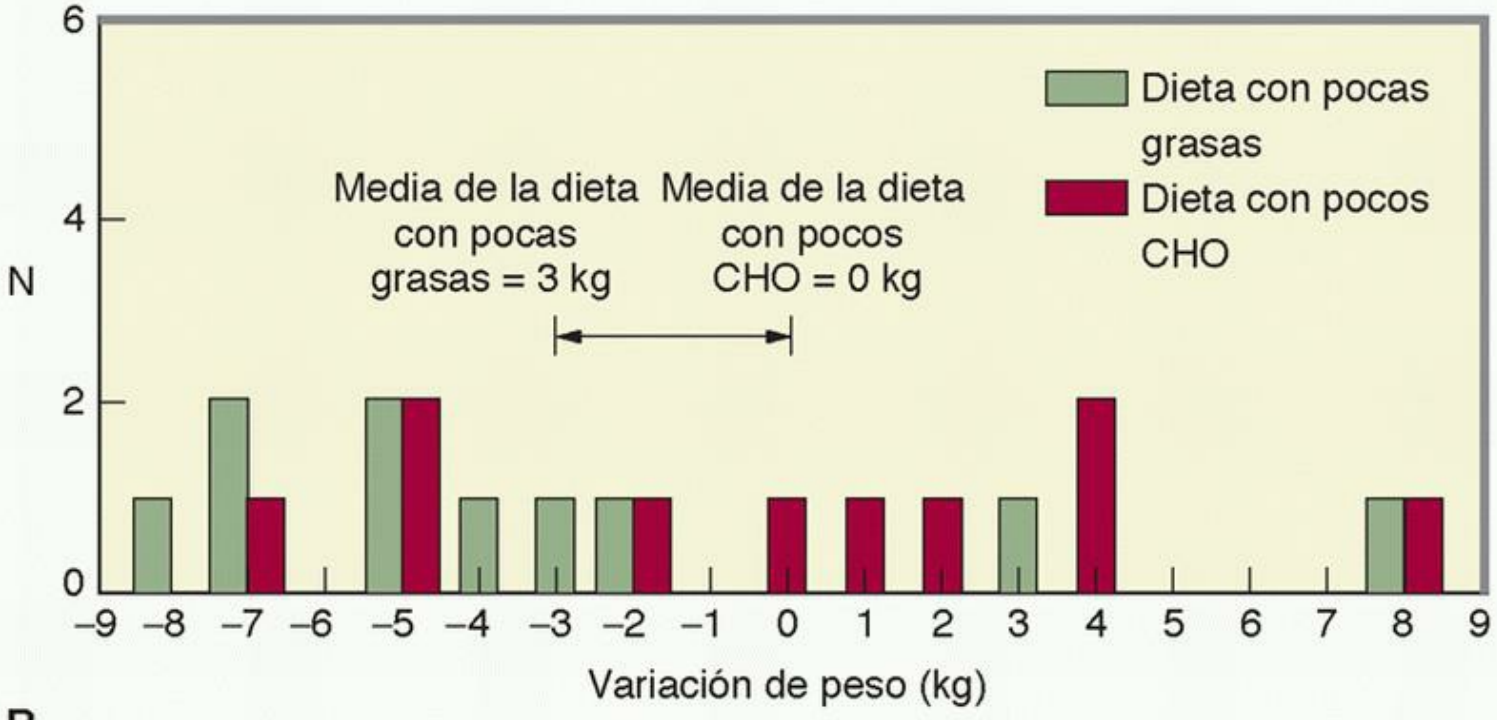
Cuando una de las variables usadas en el cálculo del tamaño de la muestra es continua (p. ej., peso corporal en la fig. 5-1), el investigador deberá calcular su variabilidad. (Para más detalles, v. la sección sobre la prueba de la *t* de Student en el cap. 6 .) En la demás situaciones, la variabilidad ya está incluida en los otros parámetros introducidos en las fórmulas y tablas del tamaño de la muestra, y no es necesario que se especifique.

Hipótesis múltiples e hipótesis posteriores

Cuando se comprueba más de una hipótesis en un estudio, especialmente si algunas de esas hipótesis se formularon después de haber analizado los datos (hipótesis posteriores), aumenta la probabilidad de que al menos una alcance la significación estadística solo por el efecto del azar. Por ejemplo, si se verifican 20 hipótesis independientes con un valor α de 0,05, hay una probabilidad elevada [64 % ($1 - 0{,}9\ 5^{20}$)] de que al menos una hipótesis sea estadísticamente significativa solo por el azar. Algunos estadísticos proponen ajustar el nivel de significación estadística cuando se verifica más de una hipótesis en un estudio. Esto mantiene la probabilidad general de aceptar cualquiera de las hipótesis alternativas, cuando todos los hallazgos se deben al azar, en el nivel especificado. Por ejemplo, los estudios genómicos que buscan una asociación entre miles de genotipos y una enfermedad necesitan usar un nivel α mucho menor de 0,05, o corren el riesgo de identificar muchas asociaciones positivas falsas.

Un método, que recibe su nombre del matemático Bonferroni, consiste en dividir el nivel de significación (es decir, 0,05) por el número de hipótesis comprobadas. Si hubiera cuatro hipótesis, por ejemplo, cada una se probaría con un nivel α de 0,0 125 (0,05/4). Esto hace que sea necesario aumentar notablemente el tamaño de la muestra por encima de lo necesario para verificar cada una de las hipótesis con un nivel α de 0,05. Así, para cualquier hipótesis particular, el abordaje de Bonferroni reduce la probabilidad de un error de tipo I a costa de aumentar la probabilidad del error de tipo II o de la necesidad de un mayor tamaño de la muestra. Si los resultados del estudio siguen siendo estadísticamente significativos después del ajuste de Bonferroni, la pérdida de potencia no es problemática. Sin embargo, es más problemático un resultado en el que se pierde la significación estadística después del ajuste de Bonferroni, lo que podría representar la imposibilidad de respaldar una asociación que estaba realmente presente en la población (error de tipo II).





▪FIGURA 5-1. A: Pérdida de peso lograda con dos dietas. Todos los que siguieron la dieta con pocas grasas perdieron de 2 a 4 kg, mientras que la variación de peso en los que siguieron la dieta baja en hidratos de carbono (CHO) variaba desde -1 a +1 kg. Como no hay superposición entre los dos grupos, es razonable inferir que la dieta baja en grasas es mejor para perder peso que la que contiene pocos hidratos de carbono (como se confirmaría mediante una prueba de la *t*, que permitió obtener un valor de $p < 0{,}0\ 001$). B: Pérdida de peso lograda con dos dietas. Existe una notable superposición en variación de peso en los dos grupos. Aunque la magnitud del efecto es la misma (3 kg) que en A, hay pocas pruebas de que una dieta sea mejor que la otra (como se confirmaría mediante una prueba de la *t*, que se asocia a un valor de *p* de 0,19).
Herramientas de imágenes

Especialmente en estos casos, decidir qué nivel de significación se va a usar depende más de la probabilidad previa de cada hipótesis que del número de hipótesis probadas, y por este motivo nuestro punto de vista general es que el uso sistemático del abordaje de Bonferroni para el estudio de múltiples hipótesis muchas veces es demasiado estricto. Existe una analogía con el uso de pruebas diagnósticas que puede ser útil (6 , 7). Cuando interpreta los resultados de una prueba diagnóstica, un médico considera la probabilidad de que el paciente

que está estudiando tenga la enfermedad en cuestión. Por ejemplo, el resultado ligeramente anómalo de una prueba en una persona sana (una concentración sérica de fosfatasa alcalina que es un 15 % mayor del límite superior de la normalidad) es probablemente una prueba falsamente positiva, que no es probable que tenga mucha importancia clínica. Igualmente, un valor de *p* de 0,05 para una hipótesis improbable es también, posiblemente, un resultado falso positivo.

Sin embargo, no es probable que una concentración de fosfatasa alcalina que sea 10 a 20 veces mayor que el límite superior normal se haya producido por casualidad (si bien pudiera ser un error de laboratorio). Así, también, no es probable que un valor de *p* muy pequeño (es decir, < 0,001) haya sucedido por casualidad (aunque pudiera deberse al sesgo). Es difícil desechar resultados de pruebas muy anómalos como positivos falsos o desechar valores de *p* muy pequeños como debidos al azar, incluso si la probabilidad previa de la enfermedad o la hipótesis era baja [2] .

Además, el número de pruebas que se habían pedido, o de hipótesis que se estudiaron, no siempre es relevante. La interpretación de una concentración de ácido úrico sérico elevada en un paciente con una articulación inflamada y dolorosa no debe depender de si el médico solicitó una sola prueba (la concentración de ácido úrico) o de si obtuvo el resultado como parte de un panel de 20 pruebas. Igualmente, cuando se interpreta el valor de *p* para verificar una hipótesis de investigación que tenga sentido, no debe importar que el investigador verifique también varias hipótesis improbables. Lo que más importa es la racionalidad de la hipótesis de investigación que se está estudiando: que tiene una probabilidad previa elevada de ser correcta. (La probabilidad previa, en este método «bayesiano», suele ser una opinión subjetiva basada en datos de otras fuentes.) Las hipótesis que se formulan durante el diseño de un estudio suelen cumplir este requisito; después de todo, ¿por qué otro motivo emplearía el investigador tiempo y esfuerzo en planificar y realizar el estudio?

¿Qué ocurre con asociaciones no previstas que aparecen durante la recogida y el análisis de los resultados de un estudio? Este proceso se denomina, a veces, generación de hipótesis o, con un enunciado menos favorable, «prospección de datos» o «expedición de pesca». Las numerosas comparaciones informales que se realizan durante el análisis de datos son una forma de estudiar múltiples hipótesis. Surge un problema similar cuando se vuelven a definir variables durante el análisis de los datos, o cuando se presentan los resultados para subgrupos de la muestra. Valores de *p* significativos para hipótesis generadas por datos que no se llegaron a plantear durante el diseño del estudio se deben, a menudo, al azar. Deben contemplarse con interés, pero con escepticismo, y debe considerarse que son una fuente de posibles preguntas de investigación para estudios futuros.

A veces, no obstante, un investigador no puede especificar una hipótesis concreta por adelantado, aunque esa hipótesis parece razonable cuando llega la hora de analizar los datos. Esto podría suceder, por ejemplo, si otros autores descubren un nuevo factor de riesgo mientras se está realizando el estudio, o si el investigador no pensó en una hipótesis concreta cuando se estaba diseñando el estudio. El punto importante no es tanto si se formuló la hipótesis antes de iniciar el estudio, como si existe una probabilidad previa razonable, basada en datos de otras fuentes, de que la hipótesis sea cierta (6 , 7).

Hay algunas ventajas específicas para definir más de una hipótesis al planificar un estudio. El uso de múltiples hipótesis no relacionadas aumenta la eficacia del estudio, haciendo que sea posible responder a más preguntas con un solo esfuerzo de investigación y descubrir más de las verdaderas asociaciones que existen en la población. También puede ser una buena idea formular varias hipótesis *relacionadas;* si los hallazgos son congruentes, las conclusiones del estudio se refuerzan. En estudios realizados en pacientes con insuficiencia cardíaca, se ha observado que el uso de inhibidores de la enzima convertidora de la angiotensina es beneficioso para reducir los ingresos de causa cardíaca, la mortalidad cardiovascular y la mortalidad total. Si se hubiera estudiado solo una de estas hipótesis, las inferencias de estos estudios habrían sido menos definitivas. Suponga que cuando se analizan estas hipótesis relacionadas y preestablecidas, solo una resulta estadísticamente significativa. Entonces, el investigador debe decidir (e intentar convencer a editores y lectores) si los resultados significativos, los no significativos o ambos grupos de resultados son correctos.

Hipótesis principales y secundarias

Algunos estudios, especialmente los grandes ensayos aleatorizados, especifican algunas hipótesis como «secundarias». Esto suele ocurrir cuando existe una hipótesis principal alrededor de la cual se ha diseñado el estudio, pero los investigadores también están interesados en otras preguntas que son menos importantes. Por ejemplo, el resultado principal de un ensayo de complementación con aportes de cinc podrían ser las hospitalizaciones o las visitas al servicio de urgencias por infecciones de las vías respiratorias superiores; un resultado secundario podrían ser los días de absentismo laboral o escolar comunicados por el propio paciente. Si se está realizando el estudio para obtener la aprobación de un fármaco, el resultado principal es lo que será más importante para el organismo regulador. Formular una hipótesis secundaria por adelantado aumenta la credibilidad de los resultados cuando se estudia la hipótesis.

Una buena norma, particularmente para ensayos clínicos, es establecer por adelantado tantas hipótesis como tengan sentido, pero especificar solo una como la hipótesis principal, que puede verificarse estadísticamente sin necesidad de plantear si se debe hacer un ajuste para estudiar múltiples hipótesis. Más importante aún, tener una hipótesis principal ayuda a centrar el estudio en su objetivo principal y proporciona una base clara para el cálculo del tamaño de la muestra principal.

Muchos estadísticos y epidemiólogos están pasando del estudio de hipótesis, con su énfasis en los valores de p, a utilizar intervalos de confianza para describir la precisión de los resultados del estudio (8 , 9 , 10). De hecho, algunos autores piensan que todo el proceso de basar la planificación del tamaño de la muestra en las hipótesis es erróneo, en parte porque depende de cantidades que son desconocidas (magnitud del efecto) o arbitrarias (α y β) (11). Sin embargo, el abordaje que hemos asumido es práctico y sigue siendo la norma en la planificación de la investigación clínica.

Volver al principio

▪ RESUMEN

- La planificación del tamaño de la muestra es una parte importante del diseño de los estudios analíticos y descriptivos. El tamaño de la muestra debe calcularse en fases tempranas del proceso de desarrollar el diseño de la investigación, de modo que puedan hacerse modificaciones adecuadas.

- Los estudios analíticos y experimentales necesitan una hipótesis que especifique, para las posteriores pruebas estadísticas, la asociación prevista entre las principales variables predictivas y de respuesta. Los estudios puramente descriptivos, que carecen de estrategia comparativa, no necesitan una hipótesis.

- Las buenas hipótesis son específicas sobre el modo en que se muestreará a la población y se medirán las variables, sencillas (solo existe una variable predictiva y una variable de respuesta) y formuladas por adelantado.

- La hipótesis nula, que propone que la variable predictiva no está asociada con las variables de respuesta, constituye la base de las pruebas de significación estadística. La hipótesis alternativa propone que sí están asociadas. Las pruebas estadísticas intentan rechazar la hipótesis nula de ausencia de asociación a favor de la hipótesis alternativa de que sí existe una asociación.

- Una hipótesis alternativa es unilateral (solo se estudiará una dirección de asociación) o bilateral (se analizarán ambas direcciones). Las hipótesis unilaterales solo deben usarse en circunstancias no habituales, cuando solo una dirección de la asociación es clínicamente o biológicamente significativa.

- En experimentos y estudios analíticos, el tamaño de la muestra es un cálculo del número de participantes necesarios para detectar una asociación de una magnitud del efecto y variabilidad determinadas con una probabilidad especificada de cometer errores de tipo I (positivos falsos) y de tipo II (negativos falsos). La probabilidad máxima de cometer un error de tipo I se denomina α la de cometer un error de tipo II se denomina β. La cantidad (1 - β) es la potencia, la posibilidad de observar una asociación con una magnitud del efecto determinada o mayor en una muestra si está realmente presente en la población.

- A menudo es deseable establecer más de una hipótesis por adelantado, aunque el investigador debe especificar una sola hipótesis principal como objetivo y para el cálculo del tamaño de la muestra. La interpretación de los hallazgos tras estudiar múltiples hipótesis en la muestra, incluyendo hallazgos no previstos que surgen a partir de los datos, se basa en una opinión sobre la probabilidad previa que representan fenómenos reales en la población.

Volver al principio

[1]En ocasiones se denominan pruebas de una y de dos colas, por las *colas* (áreas de los extremos) de las distribuciones estadísticas.

[2]Una vez más, la excepción son algunos estudios genéticos en los que se pueden explorar millones o incluso miles de millones de asociaciones.

BIBLIOGRAFÍA

1. The Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. N Engl J Med 1994;330:1 029-1 035. Citado aquí

2. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. N Engl J Med 1991;324:781-788. Citado aquí

3. The Cardiac Arrhythmia Suppression Trial II Investigators. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. N Engl J Med 1992;327:227-233. Citado aquí

4. Van Walraven C, Mahon JL, Moher D, et al. Surveying physicians to determine the minimal important difference: implications for sample-size calculation. J Clin Epidemiol 1999;52:717-723. Citado aquí

5. McKeown-Eyssen GE, Tibshirani R. Implications of measurement error in exposure for the sample sizes of case-control studies. Am J Epidemiol 1994;139:415-421. Citado aquí

6. Browner WS, Newman TB. Are all significant P values created equal? The analogy between diagnostic tests and clinical research. JAMA 1987;257:2 459-2 463. Citado aquí

7. Newman TB, Kohn, MA. Evidence-based diagnosis. New York: Cambridge University Press, 2009. Chapter 11. Citado aquí

8. Daly LE. Confidence limits made easy: interval estimation using a substitution method. Am J Epidemiol 1998; 147:783-790. Citado aquí

9. Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med 1999;130:995-1 004. Citado aquí

10. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. Ann Intern Med 1999;130:1 005-1 013. Citado aquí

11. Bacchetti P. Current sample size conventions: flaws, harms, and alternatives. BMC Med. 2010;8:17. Citado aquí

CAPÍTULO 6 Cálculo de la potencia y el tamaño de la muestra: aplicaciones y ejemplos

Warren S. Browner
Thomas B. Newman
Stephen B. Hulley

En el capítulo 5 se presentaron los principios básicos para los cálculos del tamaño de la muestra. Este capítulo presenta varias técnicas «de manual» para usar esos principios en el cálculo del tamaño de la muestra necesario para un proyecto de investigación. La primera sección aborda los cálculos del tamaño de la muestra para un estudio experimental o analítico, incluyendo algunos puntos especiales que se aplican a estos estudios, como el análisis multivariado. La segunda sección considera estudios que son fundamentalmente descriptivos. Las siguientes secciones abordan estudios que tienen un tamaño de la muestra fijo, estrategias para aumentar al máximo la potencia de un estudio y el modo de calcular el tamaño de la muestra cuando parece existir información insuficiente con la que trabajar. El capítulo finaliza con los errores habituales que hay que evitar.

Al final del capítulo se ofrecen tablas y fórmulas, en los apéndices, para varios métodos básicos de calcular el tamaño de la muestra. Además, existe una calculadora en nuestra página web (www.epibiostat.ucsf.edu/dcr/), y hay muchas páginas en Internet que pueden realizar cálculos interactivos e instantáneos del tamaño de la muestra; intente buscar por «calculadora de tamaño muestral» (*sample size calculator,* en inglés). La mayor parte de los programas estadísticos también pueden calcular el tamaño de la muestra a partir de diseños de estudios habituales.

▪ TÉCNICAS PARA EL CÁLCULO DEL TAMAÑO DE LA MUESTRA EN ESTUDIOS EXPERIMENTALES Y ANALÍTICOS

Existen diversos métodos para calcular el tamaño de la muestra en un estudio experimental o analítico, si bien todos presentan determinados pasos en común:

- Establecer la hipótesis nula y una hipótesis alternativa unilateral o bilateral.
- Seleccionar una prueba estadística adecuada de la tabla 6-1 basada en el tipo de variable predictiva y variable de respuesta que se plantea en esas hipótesis.
- Escoger una magnitud del efecto razonable (y una variabilidad adecuada, si es necesario).
- Establecer α y β. Especificar un valor de α bilateral, salvo que la hipótesis alternativa sea claramente unilateral.
- Usar la tabla o fórmula adecuada del apéndice, una calculadora en línea o un paquete estadístico para calcular el tamaño de la muestra.

Incluso si hay dudas sobre el valor exacto de uno o más de los ingredientes, es importante calcular el tamaño de la muestra al principio de la fase de diseño. Esperar hasta el último minuto para prepararlo puede llevar a un brusco despertar: puede que sea necesario empezar de nuevo con nuevos ingredientes, lo que puede significar volver a diseñar el estudio completo. Esta es la razón por la que este tema se aborda al principio de este libro.

No todos los estudios analíticos se encuadran claramente en una de las tres categorías principales de cálculo del tamaño de la muestra descritos en las secciones siguientes: utilización de la prueba de la χ^2 si las variables predictiva y de respuesta son dicotómicas, utilización de la prueba de la *t* si una es dicotómica y la otra continua, y utilización del coeficiente de correlación si las dos son continuas. Algunas de las excepciones más habituales se analizan en la sección denominada «Otras consideraciones y problemas especiales» (pág. 60).

	VARIABLE DE RESPUESTA	
VARIABLE PREDICTIVA	**DICOTÓMICA**	**CONTINUA**
Dicotómica	Prueba de la χ^2 †	Prueba de la *t*
Continua	Prueba de la *t*	Coeficiente de correlación

[*] Véase en la sección «Otras consideraciones y problemas especiales» lo que se debe realizar con las variables ordinales, o si se piensa analizar los datos con otro tipo de prueba estadística.

[†] La prueba de la χ^2 es siempre bilateral; el estadístico *Z* es el equivalente unilateral.

Prueba de la *t*

La prueba de la***t*** (a veces, denominada «prueba de la *t* de Student», por el seudónimo de su creador) suele utilizarse para determinar si el valor medio de una variable continua en un grupo difiere significativamente del valor en el otro grupo. Por ejemplo, sería adecuado usar la prueba de la *t* al comparar las puntuaciones medias de depresión en pacientes tratados con dos antidepresivos diferentes, o el índice de masa corporal medio en personas que tienen y no tienen diabetes. La prueba de la *t* supone que la distribución de la variable en cada uno de los dos grupos se aproxima a una curva normal (con forma de campana). Sin embargo, la prueba de la *t* es notablemente sólida, por lo que puede usarse en casi cualquier distribución, salvo que el número de participantes sea pequeño (menor de 30 a 40) o haya valores extremos.

Aunque la prueba de la *t* habitualmente se utiliza para comparar resultados continuos, también se puede utilizar para estimar el tamaño de la muestra para un resultado dicotómico (como en los estudios de casos y testigos) si el estudio tiene una variable predictiva continua. En esta situación, la prueba de la *t* compara el valor medio de la variable predictiva en los casos con el de los testigos.

Para calcular el tamaño de la muestra para un estudio en el que los valores medios de una variable de respuesta continua se compararán mediante una prueba de la *t* (v. ejemplo 6-1), el investigador debe:

- Establecer la hipótesis nula y si la hipótesis alternativa es unilateral o bilateral.
- Calcular la magnitud del efecto (*E*) como la diferencia del valor medio de la variable continua de los grupos de estudio.
- Estimar la variabilidad como la desviación típica (*D*) de esa variable.
- Calcular la magnitud del efecto normalizada (*E/D*), definida como la magnitud del efecto dividida por la desviación típica de la variable de respuesta.
- Establecer α y β.

La magnitud del efecto y la variabilidad pueden calcularse, a menudo, a partir de estudios previos de la bibliografía y de la consulta con expertos. En ocasiones, será necesario un pequeño estudio piloto para calcular la desviación típica de la variable (v. también la sección «Cómo calcular el tamaño de la muestra cuando no existe información suficiente» en la pág. 70). Cuando la variable de respuesta es el cambio de una medición continua (p. ej., cambio de peso durante un estudio), el investigador debe usar la desviación típica del *cambio* de esa variable (no la desviación típica de la propia variable) en los cálculos del tamaño de la muestra. La desviación típica del cambio de una variable suele ser menor que la desviación típica de la variable; por tanto, el tamaño de la muestra también será menor.

En ocasiones, un investigador no puede obtener ninguna información significativa sobre la desviación típica de una variable. En esta situación, merece la pena utilizar una cantidad denominada magnitud del efecto normalizada, que es una cantidad adimensional que posibilita la estimación del tamaño de la muestra; también simplifica las comparaciones entre las magnitudes del efecto de diferentes variables. Por ejemplo, una diferencia de 10 mg/dl de la concentración sérica de colesterol, que tiene una desviación típica poblacional de aproximadamente 40 mg/dl, sería igual a una magnitud del efecto normalizada de 0,25. Cuanto mayor sea la magnitud del efecto normalizada, menor será el tamaño de muestra necesario. En la mayor parte de los estudios, la magnitud del efecto normalizada será >0,1. Magnitudes más pequeñas son difíciles de detectar (se necesitan tamaños de muestra muy grandes) y no suelen ser muy importantes clínicamente.

EJEMPLO 6-1. Tamaño de la muestra cuando se usa la prueba de la

Problema: la pregunta de la investigación es si existe una diferencia de la eficacia del salbutamol y el bromuro de ipratropio para el tratamiento del asma. El investigador planifica un estudio aleatorizado del efecto de estos fármacos sobre el FEV_1 (volumen espiratorio máximo en el primer segundo) tras 2 semanas de tratamiento. En un estudio anterior se vio que el FEV_1 medio en personas con asma tratado era de 2,0 l, con una desviación típica de 1,0 l. El investigador desearía poder detectar una diferencia del 10 % o más del FEV_1 medio entre los dos grupos de tratamiento. ¿Cuántos pacientes se necesitan en cada grupo (salbutamol e ipratropio) con un valor de α (bilateral) = 0,05 y una potencia = 0,80?

Solución: los ingredientes para el cálculo del tamaño de la muestra son los siguientes:

1. Hipótesis nula: el FEV_1 después de 2 semanas de tratamiento es el mismo en los pacientes asmáticos tratados con salbutamol y en los tratados con ipratropio.

Hipótesis alternativa (bilateral): El FEV_1 medio después de 2 semanas de tratamiento es diferente en los pacientes asmáticos tratados con salbutamol del de los pacientes tratados con ipratropio.

2. Magnitud del efecto = 0,2 l (10 % × 2,0 l).

3. Desviación típica del FEV_1 = 1,0 l.

4. Magnitud del efecto normalizada = magnitud del efecto/desviación típica = 0,2 l/1,0 l = 0,2.

5.α (bilateral)= 0,05; β = 1-0,80 = 0,20. (Recuerde que β = 1 - potencia.)

Recorriendo en horizontal desde una magnitud del efecto normalizada de 0,20 en la columna situada más a la izquierda de la tabla 6A y en vertical desde α (bilateral) = 0,05 y β = 0,20, se necesitan 394 pacientes por grupo. Este es el número de pacientes de cada grupo que deben completar el estudio, aunque se necesitarán más para tener en cuenta los abandonos. Puede que no sea posible este tamaño de la muestra, y el investigador debería considerar de nuevo el diseño del estudio o quizás acordar poder detectar solo una magnitud del efecto mayor. Se puede ver una posible solución en la sección sobre la prueba de la *t* para muestras emparejadas (ejemplo 6-8).

En el apéndice 6A se muestran los requisitos de tamaño de la muestra para diversas combinaciones de α y β para varias magnitudes normalizadas. Para usar la tabla 6A , busque en la columna situada más a la izquierda la magnitud del efecto normalizada. A continuación, recorra la tabla hasta los valores de α y β elegidos para encontrar el tamaño de la muestra necesario por grupo. (Los números de la tabla 6A suponen que los dos grupos que se comparan tienen el mismo tamaño de muestra; use la fórmula que está bajo la tabla, un paquete estadístico o un programa interactivo en la Web si esa suposición no es cierta.)

Existe un atajo cómodo para obtener una aproximación del tamaño de la muestra usando la prueba de la *t* cuando se van a estudiar más de unos 30 participantes y se establece la potencia en 0,80 (β = 0,2) y α (bilateral) en 0,05 (1). La fórmula es:

Tamaño de la muestra (por grupo de igual tamaño) = $16/(\text{magnitud del efecto normalizada})^2$

En el ejemplo 6-1 , el cálculo del tamaño de la muestra por el atajo sería $16/0{,}2^2 = 400$ por grupo.

Prueba de la χ^2

La prueba de la ji al cuadrado (χ^2) puede usarse para comparar la proporción de participantes de cada uno de los dos grupos que tienen una variable de respuesta dicotómica. Por ejemplo, la proporción de varones que presentan cardiopatía isquémica mientras son tratados con folato puede compararse con la proporción de los que presentan cardiopatía isquémica mientras toman un placebo. Esta prueba es siempre bilateral; una prueba equivalente para hipótesis unilaterales es la prueba de la Z unilateral.

En un experimento o estudio de cohortes, la magnitud del efecto se especifica por la diferencia entre P_1, la proporción de participantes que se espera que tengan la variable de respuesta en un grupo (es decir, el riesgo del resultado), y P_2, la proporción esperada en el otro grupo. Por ejemplo, en un estudio de cohortes en el que se compara el riesgo de presentar insuficiencia renal terminal en hombres y mujeres con hipertensión, P_1 sería la proporción de hombres que presentan insuficiencia renal terminal y P_2 sería la proporción de mujeres que lo hacen. La variabilidad depende de P_1 y P_2, por lo que no es necesario especificarla.

Por el contrario, para el cálculo del tamaño de la muestra en un estudio de casos y testigos, P_1 y P_2 tienen definiciones diferentes. Se refieren a las proporciones de casos y testigos que se espera que tengan un valor particular de una variable predictiva dicotómica (p. ej., la proporción de casos de insuficiencia renal terminal que eran hombres). Así, en un estudio de casos y testigos, P_1 representa la proporción de casos que se espera que tengan una variable predictiva particular (es decir, la prevalencia de esa variable predictiva) y P_2 representa la proporción de testigos que se espera que tengan la variable predictiva.

Para estimar el tamaño de la muestra para un estudio que se analizará con la prueba de la χ^2 o la prueba de la *Z* para comparar dos proporciones (ejemplo 6.2), el investigador debe:

- Establecer la hipótesis nula y decidir si la hipótesis alternativa debe ser unilateral o bilateral.

- Calcular la magnitud del efecto y la variabilidad desde el punto de vista de P_1, la proporción con la variable de respuesta en un grupo, y de P_2, la proporción con la variable de respuesta en el otro grupo.

- Establecer α y β.

En el apéndice 6B se muestran los requisitos de tamaño de muestra para varias combinaciones de α y β y un intervalo de valores de P_1 y P_2. Para calcular el tamaño de la muestra, busque en la columna situada más a la izquierda de las tablas 6B-1 o 6B-2 el valor más pequeño de P_1 y P_2 (si es necesario, redondee al 0,05 más cercano). A continuación, recorra la tabla buscando la diferencia entre P_1 y P_2. Basándose en los valores escogidos para α y β la tabla proporciona el tamaño de la muestra necesario por grupo.

A menudo, el investigador especifica la magnitud del efecto desde el punto de vista del riesgo relativo (cociente de riesgo) de la variable de respuesta en dos grupos de participantes. Por ejemplo, un investigador podría estudiar si las mujeres que toman anticonceptivos orales tienen al menos el doble de probabilidad de sufrir un infarto de miocardio que las que no los usan. En un estudio de cohortes (o experimental) es directo transformar el riesgo relativo en las dos proporciones (P_1 y P_2) y viceversa, ya que el riesgo relativo es P_1 dividido por P_2 (o viceversa).

EJEMPLO 6-2. Cálculo del tamaño de la muestra cuando se usa la prueba de la χ^2

Problema: la pregunta de la investigación es si las personas que practican Tai Chi tienen menor riesgo de presentar dolor de espalda que las que corren para hacer ejercicio. Una revisión de la bibliografía sugiere que el riesgo de dolor de espalda a los 2 años es de aproximadamente 0,30 en corredores. El investigador espera demostrar que el Tai Chi reduce el riesgo en, al menos, 0,10. Con un valor α (bilateral) = 0,05 y una potencia = 0,80, ¿a cuántos participantes se deberá estudiar para determinar si la incidencia a los 2 años de presentar dolor de espalda es de 0,20 (o menos) en los que hacen Tai Chi?

Solución: los ingredientes para el cálculo del tamaño de la muestra son los siguientes:

1. Hipótesis nula: la incidencia de dolor de espalda es la misma en las personas que corren y en las que practican Tai Chi.

Hipótesis alternativa (bilateral): la incidencia de dolor de espalda es diferente en las personas que corren y en las que practican Tai Chi.

2.P_2 (incidencia en los que corren) = 0,30; P_1 (incidencia en los que practican Tai Chi) = 0,20. El menor de estos valores es 0,20, y la diferencia entre ellos (P_1- P_2) es 0,10.

3.α (bilateral) = 0,05; β = 1 - 0,80 = 0,20.

Recorriendo transversalmente desde 0,20 en la columna situada más a la izquierda de la tabla 6B-1 , y descendiendo desde una diferencia esperada de 0,10, el número medio para α (bilateral) = 0,05 y β = 0,20 es el tamaño de la muestra necesario de 313 corredores y 313 personas que realizan Tai Chi que deben participar en el estudio.

Sin embargo, en un estudio de casos y testigos, la situación es algo más compleja, porque el riesgo relativo debe aproximarse mediante la razón de posibilidades (RP):

El investigador debe especificar la razón de posibilidades (RP) y P_2 (la proporción de testigos expuestos a la variable predictiva). Por lo tanto, P_1 (la proporción de casos expuestos a la variable predictiva) es:

Por ejemplo, si el investigador espera que el 10 % de las testigos estarán expuestas a los anticonceptivos orales (P_2 = 0,1) y desea detectar una razón de posibilidades de 3 asociada a la exposición, entonces:

Coeficiente de correlación

Aunque el coeficiente de correlación***(r)*** no se usa habitualmente en el cálculo del tamaño de la muestra, puede ser útil cuando tanto la variable predictiva como la de respuesta son continuas. Este coeficiente es una medida de la solidez de la asociación lineal entre las dos variables. Varía entre -1 y +1. Los valores negativos indican que, a medida que una variable aumenta, la otra disminuye (como la

concentración de plomo en sangre y el CI en los niños). Cuanto más próximo está el valor de *r* a 1, más fuerte es la asociación; cuanto más próximo está a cero, más débil es la asociación. La altura y el peso en los adultos, por ejemplo, están muy relacionadas en algunas poblaciones, con $r \approx 0,9$. Estos valores altos, sin embargo, no son frecuentes; muchas asociaciones biológicas tienen coeficientes de correlación mucho más pequeños.

Los coeficientes de correlación son frecuentes en algunos campos de la investigación clínica, como la medicina conductista, pero su utilización para calcular el tamaño de la muestra presenta un inconveniente: los coeficientes de correlación tienen poco significado intuitivo. Cuando se eleva al cuadrado (r^2), un coeficiente de correlación representa la proporción de la dispersión (varianza) de una variable de respuesta que se debe a su asociación lineal con una variable predictiva, y viceversa. Este es el motivo por el que valores pequeños de *r*, como los que son $\leq 0,3$, pueden ser estadísticamente significativos si la muestra es lo suficientemente grande sin ser muy significativa clínicamente o científicamente, ya que «explican» a lo sumo el 9 % de la varianza.

Una forma alternativa (a menudo, preferida) de calcular el tamaño de una muestra para un estudio en el que las variables predictiva y de respuesta son ambas continuas es dicotomizar una de las dos variables (p. ej., utilizando como referencia la mediana) y usar, en su lugar, los cálculos de la prueba de la *t*. Esto tiene la ventaja de expresar la magnitud del efecto como una diferencia entre dos grupos (la interpretación de coeficientes de correlación, que no comunica la magnitud del efecto, es más indefinida). Para calcular el tamaño de la muestra para un estudio que se analizará con un coeficiente de correlación (ejemplo 6.3), el investigador debe:

- Establecer la hipótesis nula, y decidir si la hipótesis alternativa es unilateral o bilateral.
- Calcular la magnitud del efecto como el valor absoluto del coeficiente de correlación (*r*) más pequeño que al investigador le gustaría poder detectar. (La variabilidad es una función de *r* y está ya incluida en la tabla y en la fórmula.)
- Establecer α y β.

En el apéndice 6C, busque en la columna situada más a la izquierda de la tabla 6C la magnitud del efecto (*r*). A continuación, recorra la tabla hasta los valores escogidos de α y β que proporcionan el tamaño total requerido de la muestra. La tabla 6C proporciona el tamaño de muestra adecuado cuando el investigador desea rechazar la hipótesis nula de que no existe asociación entre las variables predictiva y de respuesta (es decir, $r = 0$). Si el investigador desea determinar si el coeficiente de correlación del estudio difiere de un valor distinto a cero (p. ej., $r = 0,4$), debe recurrir al texto que se encuentra bajo la tabla 6C para buscar el método adecuado.

EJEMPLO 6-3. Cálculo del tamaño de la muestra cuando se usa el coeficiente de correlación en un estudio transversal

Problema: la pregunta de la investigación es si las concentraciones urinarias de cotinina (una medida de la intensidad del consumo actual de cigarrillos) se relacionan con la densidad ósea en los fumadores. En un estudio previo se observó una relación pequeña ($r = -0,3$) entre el consumo comunicado (en cigarrillos/día) y la densidad ósea (en g/cm^3); el investigador prevé que las concentraciones urinarias de cotinina tendrán, al menos, relación). ¿Cuántos fumadores se necesitarán, con α (bilateral) = 0,05 y $\beta = 0,10$?

Solución: los ingredientes para el cálculo del tamaño de la muestra son los siguientes:

1. Hipótesis nula: no existe relación entre la concentración urinaria de cotinina y la densidad ósea en los fumadores.

Hipótesis alternativa: existe relación entre la concentración urinaria de cotinina y la densidad ósea en los fumadores.

2. Magnitud del efecto (r) = $| -0,3| = 0,3$.

3.α (bilateral) = 0,05; $\beta = 0,10$.

Usando la tabla 6C , leyendo horizontalmente desde $r = 0,30$, en la columna situada más a la izquierda, y descendiendo desde α (bilateral) = 0,05 y $\beta = 0,10$, se necesitarán 113 fumadores.

Volver al principio

▪ OTRAS CONSIDERACIONES Y ASPECTOS ESPECIALES

Abandonos

Cada unidad de muestreo debe estar disponible para el análisis; los participantes que intervienen en un estudio, pero en los no se puede averiguar su estado de respuesta (como los abandonos), no cuentan en el tamaño de la muestra. Si el investigador adelanta que alguno de

los participantes no estará disponible para el seguimiento (como ocurre con mucha frecuencia), debe estimar la proporción que se perderá e incrementar, en consecuencia, el tamaño de la muestra incluida. Si, por ejemplo, el investigador calcula que se perderá el 20 % de su muestra para el seguimiento, se deberá aumentar el tamaño de la muestra según un factor de (1/[1 - 0,20)], o 1,25.

Variables categóricas

Aunque hay motivos matemáticos por los que puede no ser adecuado estimar el tamaño de una muestra para variables ordinales utilizando una prueba, en la práctica, las variables ordinales muchas veces se pueden tratar como variables continuas, especialmente si el número de categorías es relativamente grande (seis o más) y el promedio de los valores de la variable tiene sentido.

En otras situaciones, la mejor estrategia es cambiar ligeramente la hipótesis de la investigación, dicotomizando la variable categórica. A modo de ejemplo, suponga que un investigador está estudiando si hablar inglés como segundo idioma se asocia al número de veces que los pacientes diabéticos consultan con el podólogo en 1 año. El número de visitas está distribuido irregularmente: muchas personas no acudirán nunca, algunas acudirán una vez y solo unas pocas acudirán dos o más veces. En esta situación, el investigador podría calcular el tamaño de la muestra como si la variable fuera dicotómica (ninguna visita frente a una o más visitas).

Análisis de supervivencia

Cuando un investigador desea comparar la supervivencia o cualquier otro dato de tiempo hasta un episodio, como cuál de los tratamientos es más eficaz en la prolongación de la vida en mujeres con cáncer de mama avanzado, el análisis de supervivencia será una técnica adecuada para analizar los datos (2 , 3). Aunque la variable de respuesta, como los meses de supervivencia, *parece* ser continua, la prueba de la *t* no es adecuada, porque lo que se está evaluando en realidad no es el tiempo (una variable continua), sino la proporción de pacientes (una variable dicotómica) que sigue con vida en cada momento a lo largo del tiempo. De manera similar, un investigador podría estar comparando la incidencia de aparición del resultado (por cada 100 personas/año de seguimiento) en dos grupos. Se puede realizar una aproximación razonable simplemente estimando la proporción de pacientes que se espera que lleguen a tener ese resultado en los dos grupos y calculando el tamaño de la muestra con la prueba de la χ^2. Sin embargo, si se espera que el resultado se produzca en la mayoría de los pacientes, como la muerte en un estudio de cáncer de mama avanzado, una estrategia mejor (porque minimiza el tamaño de la muestra total) es estimar el tamaño de la muestra basado en las proporciones de pacientes de cada grupo que se espera que tengan el resultado en un momento del seguimiento en el que se haya producido aproximadamente la mitad de los resultados totales. Por ejemplo, en un estudio en el que se compara la supervivencia sin enfermedad recurrente en pacientes con cáncer de mama tratadas con un tratamiento estándar o con otro experimental, en el que se espera que hayan muerto aproximadamente el 60 % de las pacientes del grupo de tratamiento estándar a los 2 años, en comparación con el 40 % de las que reciban el tratamiento experimental, el tamaño de la muestra se puede estimar utilizando la «supervivencia a los 2 años» como resultado dicotómico.

Muestras de conglomerados

Algunos diseños de investigación conllevan el uso de muestras de conglomerados, en las que los participantes se muestrean por grupos (cap. 11). Considere, por ejemplo, un estudio de si una intervención educativa continua para médicos mejora la tasa de abandono del tabaquismo entre sus pacientes. Suponga que se asignan aleatoriamente 20 consultas de médicos al grupo que recibe la intervención, y 20 consultas se asignan a un grupo de control. Un año después, los investigadores planean revisar las historias de una muestra aleatoria de 50 pacientes de cada consulta que eran fumadores al principio para determinar cuántos han dejado de fumar. ¿Equivale el tamaño de la muestra a 40 (el número de médicos) o a 2 000 (el número de pacientes)? La respuesta, que se encuentra en algún punto entre esos dos extremos, depende de lo similares que sean los pacientes de la consulta de un médico (en cuanto a su probabilidad de dejar de fumar) en comparación con la similitud entre todos los pacientes. Para calcular esta cantidad, a menudo se necesita conseguir datos piloto, salvo que otro investigador haya realizado anteriormente un estudio similar. Existen varias técnicas para calcular el tamaño de la muestra requerido para un estudio que utiliza muestras de conglomerados (4 , 5 , 6 , 7), pero suelen precisar la ayuda de un estadístico.

Emparejamiento

Por diversas razones, un investigador puede escoger usar un diseño emparejado (cap. 9). Sin embargo, las técnicas de este capítulo, que ignoran cualquier emparejamiento, ofrecen estimaciones razonables del tamaño de la muestra necesario, salvo que la exposición (en estudios de casos y testigos emparejados) o el resultado (en estudios de cohortes emparejados) tenga una correlación elevada con la correspondiente variable. Pueden realizarse cálculos más precisos, lo que precisa que el investigador especifique la correlación entre las exposiciones o los resultados en pares emparejados, usando abordajes normalizados (8), programas estadísticos o programas interactivos de la red.

Ajuste multivariable y otros análisis estadísticos especiales

Al diseñar un estudio de observación, un investigador puede decidir que una o más de las variables introducirán confusión en la

asociación entre la variable predictiva y la de respuesta (cap. 9), y planificar utilizar técnicas estadísticas para ajustar estos factores de confusión al analizar los resultados. Cuando se va a incluir este ajuste en la verificación de la hipótesis primaria, el tamaño de muestra calculado necesita tener esto en cuenta.

Los métodos analíticos que corrigen variables de confusión a menudo aumentan el tamaño de muestra requerido (9 , 10). La magnitud de este aumento depende de varios factores, como la prevalencia del factor de confusión, la solidez de la asociación entre la variable predictiva y la de confusión, y la solidez de la asociación entre la variable de confusión y la de respuesta. Estos efectos son complejos, y no existen reglas generales que abarquen todas las situaciones.

Los estadísticos han desarrollados métodos multivariados, como la regresión lineal y la regresión logística, que permiten al investigador corregir las variables de confusión. Una técnica estadística muy utilizada, el análisis de riesgos proporcionales de Cox, puede corregir tanto variables de confusión como diferencias en la duración del seguimiento. Si va a utilizarse una de estas técnicas para analizar los datos, existen métodos correspondientes para calcular el tamaño de muestra necesario (3 , 11 , 12 , 13 , 14). También se dispone de técnicas para calcular el tamaño de la muestra para otros diseños, como estudios de posibles factores de riesgo genético o de genes candidatos (15 , 16 , 17), estudios económicos (18 , 19 , 20), estudios de dosis y respuesta (21) o estudios en los que intervienen más de dos grupos (22). De nuevo, Internet es un recurso útil para estos métodos más sofisticados (p. ej., buscar *sample size* y *logistic regression*, tamaño de la muestra y regresión logística).

Suele ser más fácil, al menos para investigadores principiantes, calcular el tamaño de la muestra suponiendo un método de análisis más sencillo, como la prueba de la χ^2 o la prueba de la *t*. Suponga, por ejemplo, que un investigador está pensando en un estudio de casos y testigos sobre si la concentración de colesterol sérica (una variable continua) está asociada con la aparición de tumores cerebrales (una variable dicotómica). Incluso si el plan final es analizar los datos con la técnica de regresión logística, puede calcularse un tamaño de muestra aproximado con la prueba de la *t*. El resultado es que los métodos simplificados suelen producir cálculos de tamaño de muestra que son similares a los generados por técnicas más sofisticadas. Sin embargo, puede que sea necesario consultar con un bioestadístico experto si se ha enviado una propuesta económica que conlleva costes importantes para lograr financiación: los revisores de la propuesta económica esperarán que se utilice un método sofisticado, aunque reconozcan que los cálculos del tamaño de la muestra se basan en supuestos sobre el riesgo de los resultados, la magnitud del efecto, etc. Conseguir que un estadístico sea quien estime el tamaño de la muestra también transmite el mensaje de que usted tiene acceso a los colaboradores que serán necesarios para manejar y analizar los datos del estudio. De hecho, un bioestadístico contribuirá de otras muchas maneras al diseño y la ejecución del estudio, aunque seguramente agradecerá trabajar con un investigador clínico que haya pensado en los problemas y haya realizado, al menos, un intento inicial de estimación del tamaño de la muestra.

Estudios de equivalencia y de ausencia de inferioridad

A veces, el objetivo de un estudio es descartar una asociación significativa entre las variables predictiva y de respuesta. En un estudio de equivalencia se verifica si un nuevo fármaco tiene la misma eficacia que un fármaco ya utilizado. Esta situación plantea un reto cuando se planifica el tamaño de la muestra, porque la magnitud del efecto deseada es cero o muy pequeña. El estudio de ausencia de inferioridad es una versión unilateral de este diseño, que analiza si el nuevo fármaco es, al menos, no mucho peor que el fármaco ya utilizado (cap. 11).

Los cálculos del tamaño de la muestra para estos diseños son complejos (23 , 24 , 25 , 26), y será útil el consejo de un estadístico con experiencia. Un método aceptable sería diseñar el estudio para que tenga una potencia elevada (es decir, 0,90 o 0,95) para rechazar la hipótesis nula cuando la magnitud del efecto es lo suficientemente pequeña como para no ser clínicamente importante (p. ej., una diferencia de 5 mg/dl en la glucemia media en ayunas). Si los resultados de este estudio con buena potencia muestran «ausencia de efecto» (el intervalo de confianza al 95 % excluye la diferencia especificada previamente de 5 mg/dl), el investigador puede estar razonablemente seguro de que los dos fármacos tienen efectos similares. Un problema con los estudios de equivalencia y de ausencia de inferioridad, sin embargo, es que la potencia adicional y la pequeña magnitud del efecto a menudo precisan un tamaño de la muestra muy grande; de los dos diseños, los estudios de ausencia de inferioridad tienen la ventaja de ser unilaterales, lo que permite un tamaño de la muestra menor o un menor valor de α.

Otro problema supone la pérdida de las salvaguardas habituales inherentes al paradigma de la hipótesis nula, lo que protege frente a los errores de tipo I (rechazar falsamente la hipótesis nula) a los estudios convencionales que comparan un fármaco activo con un placebo. El paradigma asegura que muchos problemas del diseño o la ejecución de un estudio, como usar determinaciones imprecisas o una pérdida excesiva de pacientes durante el seguimiento, hacen que sea más difícil rechazar la hipótesis nula. Los investigadores de un estudio convencional, que están intentando rechazar una hipótesis nula, tienen un importante incentivo para realizar el mejor estudio posible. Sin embargo, estas salvaguardas no se aplican a los estudios de ausencia de inferioridad, en los que el objetivo es no encontrar ninguna diferencia.

Volver al principio

▪ TÉCNICAS PARA CALCULAR EL TAMAÑO DE LA MUESTRA EN ESTUDIOS DESCRIPTIVOS

El cálculo del tamaño de la muestra en los estudios descriptivos, entre ellos los estudios de pruebas diagnósticas, se basa también en principios algo diferentes. Estos estudios no tienen variables predictivas ni de respuesta, ni comparan grupos diferentes estadísticamente, por lo que no se aplican los conceptos de potencia y de hipótesis nula y alternativa. Por el contrario, el investigador calcula estadísticos descriptivos, como medias y proporciones. A menudo, sin embargo, también se usan estudios descriptivos (¿cuál es la frecuencia de depresiones entre pacientes ancianos en una consulta médica?) para realizar preguntas analíticas (¿cuáles son las variables predictivas en estos pacientes?). En esta situación, el tamaño de la muestra debe calcularse también para el estudio analítico, para evitar el problema habitual de tener una potencia inadecuada para lo que resulta ser la pregunta de mayor interés.

Los estudios descriptivos habitualmente comunican intervalos de confianza, un intervalo de valores sobre la proporción o la media de la muestra. Un intervalo de confianza es una medida de la precisión del cálculo de una muestra. El investigador establece el nivel de confianza, como el 95 % o el 99 %. Un intervalo con un nivel de confianza mayor (99 %) es más amplio, y, por lo tanto, es más probable que incluya el valor real de la población, que un intervalo con un menor nivel de confianza (90 %).

La amplitud de un intervalo de confianza depende del tamaño de la muestra. Por ejemplo, un investigador podría desear calcular la puntuación media obtenida por un grupo de estudiantes de medicina en el *U.S. Medical Licensing Examination* a los que se enseñó a utilizar un currículo alternativo basado en Internet. De una muestra de 50 estudiantes, podría calcular que la puntuación media en la población de todos los estudiantes es 215, con un intervalo de confianza al 95 % desde 205 a 225. Un estudio más pequeño, con 20 estudiantes, por ejemplo, podría tener aproximadamente la misma puntuación media, pero casi con toda seguridad tendría un intervalo de confianza al 95 % más amplio.

Al calcular el tamaño de la muestra en estudios descriptivos, el investigador especifica el nivel y la amplitud deseados del intervalo de confianza. El tamaño de la muestra puede, entonces, determinarse a partir de tablas y fórmulas que aparecen en el apéndice.

Variables continuas

Cuando la variable de interés es continua, con frecuencia se comunica un intervalo de confianza alrededor del valor medio de esa variable. Para calcular el tamaño de la muestra para ese intervalo de confianza (ejemplo 6.4), el investigador debe:

- Calcular la desviación típica de la variable de interés.
- Especificar la precisión deseada (amplitud total) del intervalo de confianza.
- Seleccionar el nivel de confianza para el intervalo (p. ej., 95 % o 99 %).

Para usar el apéndice 6D, normalice la amplitud total del intervalo (divídala por la desviación típica de la variable) y busque en la columna situada más a la izquierda de la tabla 6D la amplitud normalizada esperada. A continuación, recorra la tabla hasta el nivel de confianza escogido para el tamaño de muestra requerido.

Variables dicotómicas

En un estudio descriptivo de una variable dicotómica, los resultados pueden expresarse como un intervalo de confianza alrededor de la proporción calculada de participantes con uno de los valores. Esto incluye estudios de la sensibilidad y la especificidad de una prueba diagnóstica, que parecen ser, a primera vista, variables continuas, pero que en realidad son dicotómicas: proporciones expresadas como porcentajes (cap. 12). Para calcular el tamaño de la muestra para ese intervalo de confianza, el investigador debe:

EJEMPLO 6-4. Cálculo del tamaño de la muestra para un estudio descriptivo de una variable continua

Problema: el investigador pretende determinar la concentración media de hemoglobina entre alumnos de tercer grado en un área urbana con un intervalo de confianza al 95 % de ± 0,3 g/dl. En un estudio anterior se observó que la desviación típica de la hemoglobina en una ciudad similar fue de 1 g/dl.

Solución: los ingredientes para calcular el tamaño de la muestra son los siguientes:

1. Desviación típica de la variable (DT) = 1 g/dl.

2. Amplitud total del intervalo = 0,6 g/dl (0,3 g/dl por encima y 0,3 g/dl por debajo). La amplitud normalizada del intervalo = amplitud

total/DT = 0,6/1 = 0,6.

3. Nivel de confianza = 95 %.

Recorriendo transversalmente desde una amplitud estandarizada de 0,6, en la columna situada más a la izquierda de la tabla 6D , y descendiendo desde el nivel de confianza del 95 %, el tamaño de muestra necesario es de 43 alumnos de tercer grado.

- Calcular la proporción esperada de la variable de interés en la población. (Si se espera que más de la mitad de la población tenga la característica, planifique el tamaño de la muestra según la proporción que se espera que no tenga la característica.)
- Especificar la precisión deseada (amplitud total) del intervalo de confianza.
- Seleccionar el nivel de confianza para el intervalo (p. ej., 95 %).

En el apéndice 6E, busque en la columna situada más a la izquierda de la tabla 6E la proporción esperada de la variable de interés. A continuación, recorra la tabla hasta la amplitud y el nivel de confianza escogidos, que proporcionan el tamaño de muestra requerido.

En el ejemplo 6-5 se muestra un cálculo del tamaño de muestra para estudiar la sensibilidad de una prueba diagnóstica, que permite obtener el número de participantes con la enfermedad necesarios. Al estudiar la especificidad de la prueba, el investigador debe calcular el número necesario de participantes que *no* tienen la enfermedad. También hay técnicas para calcular el tamaño de la muestra en estudios de curvas de eficacia diagnóstica (ROC) (27), razones de verosimilitud (28) y fiabilidad (29) (cap. 12).

EJEMPLO 6-5. Cálculo del tamaño de la muestra para un estudio descriptivo de una variable dicotómica

Problema: el investigador desea determinar la sensibilidad de una nueva prueba diagnóstica para el cáncer de páncreas. Basándose en un estudio piloto, espera que el 80 % de los pacientes con cáncer de páncreas presenten resultados positivos en la prueba. ¿Cuántos pacientes se necesitarán para calcular un intervalo de confianza al 95 % para una sensibilidad de la prueba de 0,80 ± 0,05?

Solución: los ingredientes para el cálculo del tamaño de la muestra son los siguientes:

1. Proporción esperada = 0,20. (Como 0,80 es más de la mitad, el tamaño de la muestra se calcula a partir de la proporción que se espera que tenga un resultado negativo, es decir, 0,20.)

2. Amplitud total = 0,10 (0,05 por debajo y 0,05 por encima).

3. Nivel de confianza = 95 %.

Recorriendo transversalmente desde 0,20, en la columna situada más a la izquierda de la tabla 6E , y descendiendo desde una amplitud total de 0,10, el número medio (que representa el nivel de confianza del 95 %) proporciona el tamaño de muestra necesario de 246 pacientes con cáncer de páncreas.

Volver al principio

▪ QUÉ HACER CUANDO EL TAMAÑO DE LA MUESTRA ES FIJO

Especialmente cuando se realizan análisis de datos secundarios, el tamaño de la muestra puede haberse determinado antes del diseño del estudio. Incluso cuando se diseñó un estudio desde el principio, es frecuente encontrar que el número de participantes que están disponibles o que se pueden conseguir para el estudio es bajo. De hecho, la mayoría de los investigadores, si son sinceros, reconocerán que muchas veces «trabajan hacia atrás» desde un tamaño de la muestra fijo realista para determinar la magnitud del efecto que podrán detectar con una potencia razonable. Este es, en parte, el motivo por el que es estúpido tratar una estimación del tamaño de la muestra como si estuviera grabada en piedra.

Cuando un investigador debe trabajar hacia atrás desde el tamaño de la muestra fijo (ejemplo 6-6), calcula la magnitud del efecto que puede detectarse con una potencia determinada (generalmente el 80 %). Con menos frecuencia, estima la potencia para detectar un efecto concreto. El investigador puede usar las tablas de tamaños de las muestras que aparecen en los apéndices del capítulo, interpolando cuando sea necesario, o usar las fórmulas de tamaños de las muestras de los apéndices para calcular la magnitud del efecto.

Una norma general es que un estudio debe tener una potencia del 80 % o mayor para detectar una magnitud del efecto razonable. No hay nada mágico en el 80 %: en ocasiones, un investigador tiene suerte y encuentra un resultado estadísticamente significativo incluso aunque tenga una potencia baja para realizarlo (incluso una potencia tan baja como el 50 % ofrece una probabilidad del 50-50 de

observar en la muestra un efecto estadísticamente significativo que realmente está presente en la población). Por lo tanto, puede estar justificado realizar estudios que tengan una potencia menor del 80 % si el coste de hacerlo es pequeño, como cuando se realiza un análisis de datos que ya se han recogido. Y hay algunos estudios (p. ej., uno en el que se demostró que un nuevo tratamiento reduce la presión arterial pulmonar en más del 50 % en pacientes con hipertensión pulmonar refractaria de larga evolución) en los que un tamaño de la muestra de dos o tres participantes sería suficiente para indicar que está justificado un estudio adicional (sobre seguridad y efectos sobre los resultados clínicos).

El investigador no debe olvidar, sin embargo, que debería enfrentarse a la dificultad de interpretar (y publicar) un estudio en el que no se encontró ninguna asociación a causa de una potencia insuficiente; los amplios intervalos de confianza revelarán la posibilidad de un notable efecto en la población de la cual se extrajo la pequeña muestra del estudio. También es importante comprender que un estudio con «potencia insuficiente» que tuvo «suerte» y ofreció un resultado estadísticamente significativo puede ser criticado porque los revisores son escépticos en relación con si el investigador realmente quería buscar esa asociación particular o si estudió docenas de hipótesis y seleccionó el único resultado que tenía un valor de p significativo.

EJEMPLO 6-6. Cálculo de la magnitud del efecto detectable cuando el tamaño de la muestra es fijo

Un investigador estima que tendrá acceso a 200 nuevas madres de gemelos durante su residencia. De acuerdo con un pequeño estudio piloto, estima que aproximadamente la mitad de esas mujeres (es decir, 100) podrían estar deseosas de participar en un estudio de si un programa de meditación durante 6 semanas reduce el estrés, en comparación con un grupo testigo, que recibirá un folleto en el que se describe la relajación. Si se espera que la desviación típica de la puntuación de estrés sea de 5 puntos en los grupos testigo y de tratamiento, ¿qué diferencia podrá detectar el investigador entre los dos grupos, con α (bilateral) = 0,05 y β = 0,20?

Solución: en la tabla 6A , descendiendo desde α (bilateral) = 0,05 y β = 0,20 (la columna situada más a la derecha en la tríada media de números), se necesitan 45 pacientes por grupo para detectar una magnitud del efecto normalizada de 0,6, que es igual a 3 puntos (0,6 × 5 puntos). El investigador (que tendrá unas 50 pacientes por grupo) podrá detectar una diferencia de algo menos de 3 puntos entre los dos grupos.

Volver al principio

▪ ESTRATEGIAS PARA REDUCIR AL MÍNIMO EL TAMAÑO DE LA MUESTRA Y AUMENTAR AL MÁXIMO LA POTENCIA

Cuando el tamaño de la muestra calculado es mayor que el número de participantes que realmente pueden estudiarse, el investigador debe seguir varios pasos. En primer lugar, debe comprobarse el cálculo, ya que es fácil cometer errores. A continuación deben revisarse los «ingredientes». ¿Es la magnitud del efecto irracionalmente pequeña o la variabilidad irracionalmente grande? ¿Son α o β extremadamente pequeños, el nivel de confianza demasiado elevado o el intervalo de confianza excesivamente estrecho?

Estos ajustes técnicos pueden resultar útiles, pero es importante comprender que las pruebas estadísticas dependen, finalmente, de la información contenida en los datos. Muchos cambios en los ingredientes, como la reducción de la potencia desde el 90 % al 80 %, no mejoran la cantidad ni la calidad de los datos que se recopilarán. Hay, no obstante, varias estrategias para reducir el tamaño de la muestra necesario o para aumentar la potencia para un tamaño de la muestra determinado que realmente aumentan la información contenida en los datos recopilados. Muchas de estas estrategias conllevan modificaciones de la hipótesis de investigación; el investigador deberá considerar meticulosamente si la nueva hipótesis sigue respondiendo a la pregunta de la investigación que desea estudiar.

Usar variables continuas

Cuando las variables continuas son una opción, suelen permitir usar tamaños de la muestra más pequeños que las variables dicotómicas. La presión arterial, por ejemplo, puede expresarse en milímetros de mercurio (continua) o como la presencia o ausencia de hipertensión (dicotómica). El primer caso permite un tamaño de muestra menor para una potencia concreta, o una potencia mayor para un tamaño de muestra determinado.

EJEMPLO 6-7. Uso de variables continuas o de variables dicotómicas

Problema: considere un estudio controlado con placebo para determinar el efecto de los complementos nutricionales sobre la fuerza en ancianos que viven en residencias. Estudios anteriores han establecido que la fuerza del cuádriceps (como máximo momento de torsión en newtonmetro) está distribuida de forma aproximadamente normal, con una media de 33 N·m y una desviación típica de 10 N·m, y que alrededor del 10 % de los ancianos tienen músculos muy débiles (fuerza < 20 N·m). Se piensa que los complementos nutricionales durante 6 meses están justificados si pueden aumentar la fuerza en N·m en comparación con la dieta habitual. Puede calcularse esta

variación de la fuerza media, basándose en la distribución de la fuerza del cuádriceps en el anciano, para que corresponda a una reducción de la proporción de los ancianos que están muy débiles de alrededor del 5 %.

Un diseño podría tratar la fuerza como una variable dicotómica: muy débil frente a no muy débil. Otro podría usar toda la información contenida en la determinación y tratar la fuerza como una variable continua. ¿Cuántos participantes necesitaría cada diseño siendo α (bilateral) = 0,05 y β = 0,20?

Solución: los ingredientes para el cálculo del tamaño de la muestra usando una variable de respuesta dicotómica (muy débil frente a no muy débil) son los siguientes:

1. Hipótesis nula: la proporción de ancianos en residencias que están muy débiles (máximo momento de torsión del cuádriceps < 20 N·m) tras recibir complementos nutricionales durante 6 meses es la misma que la proporción de los que están muy débiles con una dieta normal.

Hipótesis alternativa: la proporción de ancianos de residencias que están muy débiles (máximo momento de torsión del cuádriceps < 20 N·m) tras recibir complementos nutricionales durante 6 meses difiere de la proporción de los que siguen una dieta habitual.

2.P_1 (proporción de personas muy débiles con una dieta habitual) = 0,10; P_2 (en el grupo con complementos) = 0,05. El menor de estos valores es 0,05, y la diferencia entre ellos $(P_1 - P_2)$ es 0,05.

3.α (bilateral) = 0,05; β = 0,20.

Usando la tabla 6B-1 , recorriendo transversalmente desde 0,05 en la columna situada más a la izquierda y descendiendo desde una diferencia esperada de 0,05, para encontrar el número central (para α [bilateral] = 0,05 y β = 0,20), este diseño necesitaría 473 participantes por grupo.

Los ingredientes para el cálculo del tamaño de la muestra usando una variable de respuesta continua (fuerza del cuádriceps como máximo momento de torsión) son los siguientes:

1. Hipótesis nula: la fuerza media del cuádriceps (como máximo momento de torsión en N·m) en ancianos de residencias tras recibir complementos nutricionales durante 6 meses es la misma que la fuerza media del cuádriceps en los que siguen una dieta habitual.

Hipótesis alternativa: la fuerza media del cuádriceps (como máximo momento de torsión en N·m) en ancianos de residencias tras recibir complementos nutricionales durante 6 meses difiere de la fuerza media del cuádriceps en los que siguen una dieta habitual.

2. Magnitud del efecto = 5 N·m.

3. Desviación estándar de la fuerza del cuádriceps = 10 N·m.

4. Magnitud del efecto normalizada = magnitud del efecto/desviación típica = 5 N·m/10 N·m = 0,5.

5.α (bilateral) = 0,05; β = 0,20.

Usando la tabla 6A , recorriendo transversalmente desde una magnitud del efecto normalizada de 0,50, siendo α (bilateral) = 0,05 y β = 0,20, este diseño necesitaría unos 64 participantes en cada grupo. (En este ejemplo, el atajo para el cálculo del tamaño de la muestra de la página 57 de 16/[magnitud del efecto normalizada]^2, o 16/[0,5]^2 da el mismo cálculo de 64 participantes por grupo.) La consecuencia es que el uso de una variable de respuesta continua lleva a un tamaño del efecto mucho menor.

En el ejemplo 6-7 , la variable de respuesta continua permite analizar el efecto de los complementos nutritivos sobre la fuerza muscular en los ancianos. La variable de respuesta dicotómica permite analizar sus efectos sobre la proporción de participantes que tienen al menos una mínima cantidad de fuerza, lo que puede ser una medida indirecta más válida de la posible morbilidad relacionada con las caídas.

Usar mediciones emparejadas

En algunos experimentos o estudios de cohortes con variables de respuesta continuas pueden realizarse determinaciones emparejadas (una en estado basal, otra al finalizar el estudio) en cada uno de los participantes. La variable de respuesta es el cambio producido entre estas dos determinaciones. En esta situación puede utilizarse una prueba de la *t* en las determinaciones emparejadas para comparar el valor medio de este cambio en los dos grupos. Esta técnica, a menudo, permite un menor tamaño de la muestra, porque, al comparar cada participante consigo mismo, elimina la parte intraindividual basal de la variabilidad de la variable de respuesta. Por ejemplo, el

cambio de peso con una dieta tiene menos variabilidad que el peso final, porque este último está muy relacionado con el peso inicial. El tamaño de la muestra para este tipo de prueba de la *t* se calcula del modo habitual (ejemplo 6-8), con la excepción de que la magnitud del efecto normalizada (*E/D* en la tabla 6A) es la diferencia prevista del *cambio* de la variable dividido por la desviación típica *de ese cambio*.

EJEMPLO 6-8. Uso de la prueba de la

Problema: recuerde el ejemplo 6-1 , en el que el investigador que estudia el tratamiento del asma está interesado en determinar si el salbutamol puede mejorar el FEV_1 en 200 ml en comparación con el bromuro de ipratropio. Los cálculos del tamaño de la muestra indicaron que se necesitaban 394 participantes por grupo, más de los que probablemente se dispondrá. Afortunadamente, un compañero señala que los pacientes asmáticos presentan grandes diferencias de los valores del FEV_1 antes del tratamiento. Estas diferencias entre participantes explican gran parte de la variabilidad del FEV_1 tras el tratamiento, lo que enmascara el efecto del tratamiento. El investigador sugiere el uso de una prueba de la *t* para dos muestras emparejadas para comparar los *cambios* del FEV_1 en los dos grupos. En un estudio piloto se observa que la desviación típica del cambio del FEV_1 es solo de 250 ml. ¿Cuántos participantes se necesitarían por grupo, siendo α (bilateral) = 0,05 y $\beta = 0,20$?

Solución: los ingredientes para el cálculo del tamaño de la muestra son los siguientes:

1. Hipótesis nula: el cambio del FEV_1 medio tras 2 semanas de tratamiento con salbutamol es el mismo en los pacientes asmáticos tratados con salbutamol que en los pacientes tratados con bromuro de ipratropio.

Hipótesis alternativa: el cambio en el FEV_1 tras 2 semanas de tratamiento es diferente en los pacientes asmáticos tratados con salbutamol y en los tratados con bromuro de ipratropio.

2. Magnitud del efecto = 200 ml.

3. Desviación típica de la variable de respuesta = 250 ml.

4. Magnitud del efecto normalizada = magnitud del efecto/desviación típica = 200 ml/250 ml = 0,8.

5.α (bilateral) = 0,05; $\beta = 1 - 0,80 = 0,20$.

Usando la tabla 6A , este diseño necesitaría unos 26 participantes por grupo, un tamaño de muestra mucho más razonable que el de 394 por grupo del ejemplo 6-1 . En este ejemplo, el atajo del cálculo del tamaño de la muestra de $16/(\text{magnitud del efecto normalizada})^2$, o $16/0,8^2$ proporciona un cálculo similar de 25 participantes por grupo.

Una breve nota técnica

Este capítulo se refiere siempre a pruebas de la ***t*** con dos muestras, que se usan para comparar los valores medios de variables continuas en dos grupos de participantes. Una prueba de la *t* con dos muestras puede ser no emparejada, si se está comparando la variable de respuesta entre dos grupos (ejemplo 6-1), o emparejada, si la variable de respuesta es el cambio en un par de determinaciones, por ejemplo, antes y después de una intervención (p. ej., ejemplo 6-8).

Un tercer tipo de prueba de la *t*, la prueba de la ***t*** emparejada con ***una muestra***, compara el cambio medio en un par de determinaciones en un solo grupo hasta el cambio cero. Este tipo de análisis es razonablemente habitual en diseños de series temporales (cap. 11), un método de antes y después para examinar tratamientos que son difíciles de aleatorizar (p. ej., el efecto de la histerectomía electiva, una decisión que pocas mujeres están dispuestas a dejar al resultado del lanzamiento de una moneda, sobre la calidad de vida). Sin embargo, es un diseño más débil, porque la ausencia de un grupo de comparación hace que sea difícil saber qué habría ocurrido con las participantes sin tratar. Al planear un estudio que será analizado con una prueba de la *t* emparejada con una sola muestra, el tamaño de la muestra total es tan solo la mitad del tamaño de la muestra por grupo enumerado en el apéndice 6A. Por ejemplo, para $\alpha = 0,05$ (bilateral) y $\beta = 0,2$, para detectar una diferencia de la desviación típica de 0,5 (*E/D* = 0,5), harían falta 64/2 = 32 participantes. En el apéndice 6F se presenta información adicional sobre el uso correcto y erróneo de pruebas de la *t* con una y con dos muestras.

Usar variables más precisas

Debido a que reducen la variabilidad, las variables más precisas permiten un menor tamaño de la muestra tanto en estudios analíticos como en estudios descriptivos. Incluso un pequeño cambio de precisión puede tener un notable efecto sobre el tamaño de la muestra. Por ejemplo, al usar la prueba de la *t* para calcular el tamaño de la muestra, una disminución del 20 % de la desviación típica de la variable de respuesta produce una disminución del tamaño de la muestra del 36 %. En el capítulo 4 se presentan técnicas para aumentar la precisión

de una variable, como realizar determinaciones por duplicado.

Usar tamaños de grupos desiguales

Debido a que un mismo número de participantes en cada uno de dos grupos suele proporcionar la mayor potencia para un número total determinado de estos, las tablas 6A , 6B-1 y 6B-2 de los apéndices suponen tamaños de muestra iguales en los dos grupos. A veces, sin embargo, la distribución de participantes no es igual en los dos grupos, o es más fácil o más barato reunir participantes del estudio de un grupo que de otro. Puede ocurrir, por ejemplo, que un investigador quiera calcular el tamaño de la muestra para un estudio comparando el 30 % de los integrantes de una cohorte que fuma cigarrillos con el 70 % que no lo hace. O, en un estudio de casos y testigos, el número de personas con la enfermedad puede ser pequeño, pero puede que sea posible muestrear un número más grande de testigos. En general, lo que gana la potencia cuando se aumenta el tamaño de un grupo al doble del tamaño del otro es considerable; triplicar o cuadriplicar uno de los grupos proporciona ganancias progresivamente más pequeñas. Los tamaños de la muestra en grupos desiguales pueden calcularse a partir de las fórmulas que se encuentran en el texto de los apéndices 6A y 6B, o con calculadoras del tamaño de la muestra en programas estadísticos o en Internet.

Existe un método útil (30) para calcular el tamaño de la muestra en estudios de casos y testigos de factores de riesgo y variables de respuesta dicotómicos usando c testigos por cada caso (ejemplo 6-9). Si n representa el número de casos que se habrían necesitado para un testigo por caso (con unos valores concretos de α, β, y magnitud del efecto), el número de casos aproximado (n') con cn' testigos que se necesitarán es:

$$n' = ([c + 1]/2c) \times n$$

Por ejemplo, con $c = 2$ testigos por caso, entonces $([2 + 1]/[2 \times 2]) \times n = 3/4 \times n$, y solo se necesita el 75 % más de casos. Cuando c se hace mayor, n' se acerca al 50 % de n (cuando $c = 10$, por ejemplo, $n' = 11/20 \times n$).

Usar una variable de respuesta más frecuente

Cuando se planifica un estudio con una variable de respuesta dicotómica, cuanto mayor sea la frecuencia con la que se produce el resultado, hasta una frecuencia de aproximadamente 0,5, mayor será la potencia. Por lo tanto, la modificación de la definición de un resultado es una de las mejores formas de aumentar la potencia. Si una variable de respuesta se produce más a menudo, es más probable que se detecten sus variables predictivas. De hecho, la potencia depende más del número de participantes con una variable de respuesta especificada que del número total de participantes en el estudio. Los estudios con variables de respuesta raras, como la aparición de cáncer de mama en mujeres sanas, precisan tamaños de muestra muy grandes para contar con una potencia adecuada.

EJEMPLO 6-9. Uso de múltiples testigos por caso en un estudio de casos y testigos

Problema: un investigador está estudiando si la exposición a insecticidas domésticos es un factor de riesgo de sufrir anemia aplásica. El cálculo del tamaño de la muestra original indicó que se necesitarían 25 casos, usando un control por caso. Suponga que el investigador tiene acceso solo a 18 casos. ¿Qué debería hacer el investigador?

Solución: el investigador debe considerar usar múltiples testigos por caso (después de todo, puede encontrar muchos pacientes sin anemia aplásica). Usando tres testigos por caso, por ejemplo, el número aproximado de casos que se necesitarán es $([3 + 1]/[2 \times 3]) \times 25 = 17$.

EJEMPLO 6-10. Uso de variables de respuesta más habituales

Problema: suponga que un investigador está comparando la eficacia de un antiséptico para hacer gárgaras con un placebo en la prevención de las infecciones de las vías respiratorias superiores. Sus cálculos iniciales indicaban que su muestra prevista de 200 estudiantes voluntarios era inadecuada, en parte debido a que esperaba que solo alrededor del 20 % de sus participantes tuvieran una infección de las vías respiratorias superiores durante el período de seguimiento de 3 meses. Sugiera algunos cambios en el plan del estudio.

Solución: existen dos posibles soluciones: *1)* estudiar una muestra de internos y residentes de pediatría, que es probable que presenten una incidencia mucho mayor de infecciones de vías respiratorias superiores que los estudiantes; *2)* realizar el estudio en invierno, cuando esas infecciones son más frecuentes, o *3)* realizar un seguimiento de la muestra durante un período más largo, por ejemplo, 6 o 12 meses. Todas estas soluciones conllevan la modificación de la hipótesis de la investigación, pero ninguna de ellas parece lo suficientemente grande como para afectar a la pregunta global de la investigación sobre la eficacia del antiséptico para hacer gárgaras.

Una de las mejores formas para que un resultado se produzca con más frecuencia es incluir participantes con mayor riesgo de presentar esa variable (como mujeres con antecedentes familiares de cáncer de mama).

Otras serán ampliar el período de seguimiento, de modo que exista más tiempo para acumular resultados, o ampliar la definición de lo que constituye una respuesta (p. ej., incluyendo el carcinoma *in situ* ductal). Todas estas técnicas (ejemplo 6.10), no obstante, pueden modificar la pregunta del estudio, por lo que deben usarse con precaución.

Volver al principio

▪ CÓMO CALCULAR EL TAMAÑO DE LA MUESTRA CUANDO LA INFORMACIÓN ES INSUFICIENTE

A menudo, el investigador observa que está perdiendo uno o más de los ingredientes para el cálculo del tamaño de la muestra y se siente frustrado en sus intentos para planificar su estudio. Se trata de un problema especialmente frecuente cuando el investigador usa un instrumento de su diseño (como un nuevo cuestionario en el que se compara la calidad de vida de mujeres con incontinencia urinaria de esfuerzo o con tenesmo). ¿Cómo debe decidir qué fracción de una desviación típica de las puntuaciones de su instrumento sería clínicamente significativa?

La primera estrategia será una amplia búsqueda de hallazgos anteriores y relacionados sobre el tema y sobre preguntas de investigación similares. Puede bastar con situaciones aproximadamente comparables o hallazgos mediocres o antiguos. Por ejemplo, ¿hay datos de calidad de vida en pacientes con otros problemas urológicos o con afecciones relacionadas, como una colostomía? Si no se consigue nada con la revisión bibliográfica, se debe contactar con otros investigadores y pedir su opinión sobre qué se puede esperar, y preguntar si conocen algún resultado publicado que pueda ser relevante.

Si sigue sin disponerse de información, se puede considerar la realización de un estudio piloto o la obtención de una serie de datos para un análisis secundario con el fin de obtener los ingredientes perdidos antes de emprender el estudio principal. En realidad, es muy recomendable un estudio piloto en casi todos los estudios en los que intervienen nuevos instrumentos, métodos de medida o estrategias de inclusión de pacientes. Ahorra tiempo al final, facilitando a los investigadores la realización de un trabajo mucho mejor al planear el estudio principal. Los estudios pilotos son útiles para calcular la desviación típica de una determinación o la proporción de participantes con una característica concreta. Sin embargo, una alternativa es reconocer que para las variables continuas con una distribución general en forma de campana puede calcularse la desviación típica como un cuarto de la diferencia entre los extremos mayor y menor de un intervalo de valores que se producen habitualmente, ignorando los valores extremos. Por ejemplo, si es probable que la mayoría de los participantes tengan una concentración de sodio sérico entre 135 y 143 mEq/l, la desviación típica del sodio sérico es de unos 2mEq/l (1/4 × 8 mEq/l).

Otra estrategia cuando hay dudas sobre la media y la desviación típica de una variable continua o categórica es dicotomizar la variable. Las categorías se pueden agregar en dos grupos, y las variables continuas se pueden separar en la media o la mediana. Por ejemplo, dividir la calidad de vida en «mejor que la mediana» y «la mediana o menos» evita tener que estimar la desviación típica de la muestra, aunque todavía se tiene que estimar la proporción de pacientes que estarían por encima de la media general en cada uno de los dos grupos en estudio. Después, se puede utilizar la prueba de la χ^2 para hacer una estimación razonable, aunque algo elevada, del tamaño de la muestra.

Sin embargo, muchas veces el investigador debe elegir la magnitud del efecto detectable de acuerdo con un valor que considera que tiene significado clínico. En esa situación, el investigador debe comentar su elección con compañeros que conozcan el tema. Por ejemplo, suponga que un investigador está estudiando un nuevo tratamiento invasivo para la gastroparesia refractaria grave, enfermedad de la que, como mucho, el 5 % de los pacientes mejoran espontáneamente. Si se demuestra que el tratamiento es eficaz, sus colegas digestólogos indicarían que estarían deseosos de tratar hasta a cinco pacientes para obtener una mejoría mantenida en tan solo uno de ellos (como el tratamiento tiene efectos adversos graves y es costoso, no piensan que el número sea mayor de cinco). Un número que es necesario tratar (NNT) de 5 corresponde a una diferencia de riesgos del 20 % (NNT = 1/diferencia de riesgos), por lo que el investigador debe estimar el tamaño de la muestra basado en una comparación de $P_1 = 5$ % con $P_2 = 25$ % (es decir, 59 pacientes por grupo con una potencia de 0,8 y un valor de α bilateral de 0,05).

Si todo esto falla, el investigador deberá hacer un supuesto mejorado sobre los valores probables de los ingredientes perdidos. El proceso de pensar sobre el problema e imaginar los hallazgos dará lugar, a menudo, a un cálculo razonable, y de eso se trata en la planificación del tamaño de la muestra. Esta suele ser una mejor opción que solo decidir, en ausencia de cualquier justificación, diseñar el estudio para tener una potencia del 80 % con un valor a bilateral de 0,05 a fin de detectar una magnitud del efecto normalizada de, por ejemplo, 0,5 entre los dos grupos ($n = 64$, por grupo). Muy pocos revisores de subvenciones aceptarán ese tipo de decisión arbitraria.

Volver al principio

▪ ERRORES HABITUALES QUE HAY QUE EVITAR

Muchos investigadores inexpertos (¡y algunos con experiencia!) comenten errores al planificar el tamaño de la muestra. He aquí algunos de los más frecuentes:

- Un error frecuente es calcular tarde el tamaño de la muestra durante el diseño del estudio. Hágalo al principio del proceso, cuando todavía pueden hacerse cambios fundamentales.
- Las variables dicotómicas pueden parecer continuas cuando se expresan como un porcentaje o una tasa. Por ejemplo, la situación vital (vivo o muerto) podría interpretarse erróneamente como continua cuando se expresa como porcentaje de vivos. Igualmente, en el análisis de supervivencia, en el que no todos los participantes mueren, un resultado dicotómico puede parecer continuo (p. ej., mediana de supervivencia en meses). Por todo esto, la propia variable de respuesta es realmente dicotómica (una proporción), y el método sencillo adecuado al planificar el tamaño de la muestra sería la prueba de la χ^2.
- El tamaño de la muestra calcula el número de participantes con datos de respuesta, y no el número que se necesita que participe. El investigador debe planificar siempre pensando en los abandonos y en los participantes con datos ausentes.
- Las tablas que se ofrecen al final del capítulo suponen que los dos grupos que se están estudiando tienen tamaños de muestra iguales. A menudo, esto no es así; por ejemplo, un estudio de cohortes sobre si el uso de complementos vitamínicos reduce el riesgo de sufrir quemaduras solares no incluiría, probablemente, el mismo número de personas que toman o no toman vitaminas. Si los tamaños de las muestras no son iguales, deben usarse las fórmulas que siguen a las tablas o calculadoras de Internet, o programas estadísticos.
- Cuando se utiliza la prueba de la *t* para calcular el tamaño de la muestra, la desviación típica de la variable de respuesta es un factor fundamental. Por lo tanto, si la respuesta es el cambio de una variable continua, el investigador debe utilizar la desviación típica de ese cambio, en lugar de la desviación típica de la propia variable.
- Tenga en mente la posibilidad de que haya conglomerados de datos. Si parece haber dos «niveles» de tamaño de muestra (p. ej., uno para médicos y otro para pacientes), es probable que haya agrupamiento, y no se aplicarán las tablas de los apéndices.
- Si encuentra que tiene dificultad para estimar el tamaño de la muestra para su estudio, asegúrese de que su hipótesis de investigación cumpla los criterios que se han discutido previamente en este capítulo (sencillo, específico y formulado por anticipado).

Volver al principio

▪ RESUMEN

- Al calcular el tamaño de la muestra para un estudio analítico, deben darse los siguientes pasos:
  - establecer las hipótesis nula y alternativa, especificando el número de lados;
  - seleccionar una prueba estadística que pueda usarse para analizar los datos, según los tipos de variables predictivas y de respuesta (prueba de la χ^2 si las dos son dicotómicas, prueba de la ***t*** si una es dicotómica y la otra continua, y coeficiente de correlación si las dos son continuas);
  - calcular la magnitud del efecto (y su variabilidad, si es necesario); y
  - especificar valores adecuados de α y β según la importancia de evitar los errores de tipo I y de tipo II.
- Otras consideraciones al calcular el tamaño de la muestra para estudios analíticos son el ajuste para tener en cuenta los posibles abandonos, y estrategias para abordar variables categóricas, análisis de supervivencia, muestras de conglomerados, ajuste multivariado y abordajes estadísticos especiales para los estudios de equivalencia y ausencia de inferioridad.
- Los pasos para calcular el tamaño de la muestra para estudios descriptivos, que carecen de hipótesis, son: *a)* calcular la proporción de participantes con una respuesta dicotómica o la desviación típica de una respuesta continua; *b)* especificar la precisión deseada (amplitud del intervalo de confianza), y *c)* especificar el nivel de confianza (p. ej., 95 %).
- Cuando el tamaño de la muestra está predeterminado, el investigador puede trabajar retrospectivamente para calcular la magnitud del efecto detectable o, con menos frecuencia, la potencia del estudio.

- Las estrategias para reducir al mínimo el tamaño de muestra comprenden: usar variables continuas, determinaciones más precisas, determinaciones emparejadas y más respuestas frecuentes, además de aumentar el número de testigos por caso en los estudios de casos y testigos.
- Cuando no parece haber suficiente información para calcular el tamaño de la muestra, el investigador debe revisar la bibliografía en áreas relacionadas y consultar con compañeros para poder elegir un tamaño de la muestra que sea clínicamente significativo.
- Los errores que se deben evitar incluyen estimación del tamaño de la muestra demasiado tarde, interpretación errónea de las proporciones expresadas como porcentajes, no tener en consideración los sujetos y los datos ausentes, y no abordar adecuadamente los datos agrupados y pareados.

Volver al principio

APÉNDICE 6ATamaño de muestra necesario por grupo al usar la prueba de la *t* para comparar medias de variables continuas
NA

UNILATERAL	α=	**0,005**			**0,025**			**0,05**		
BILATERAL	α=	**0,01**			**0,05**			**0,10**		
E/D[1]	β=	**0,05**	**0,10**	**0,20**	**0,05**	**0,10**	**0,20**	**0,05**	**0,10**	**0,20**
0,10		3.565	2.978	2.338	2.600	2.103	1.571	2.166	1.714	1.238
0,15		1.586	1.325	1.040	1.157	935	699	963	762	551
0,20		893	746	586	651	527	394	542	429	310
0,25		572	478	376	417	338	253	347	275	199
0,30		398	333	262	290	235	176	242	191	139
0,40		225	188	148	164	133	100	136	108	78
0,50		145	121	96	105	86	64	88	70	51
0,60		101	85	67	74	60	45	61	49	36
0,70		75	63	50	55	44	34	45	36	26
0,80		58	49	39	42	34	26	35	28	21
0,90		46	39	32	34	27	21	28	22	16

1,00	38	32	26	27	23	17	23	18	14

[1] *E/D* es la magnitud del efecto normalizada, calculada como *E* (magnitud del efecto esperada) dividida por *D* (desviación típica de la variable de respuesta). Para calcular el tamaño de la muestra, recorra transversalmente desde la *magnitud del efecto normalizada* y descienda desde los valores especificados de α y β para el tamaño de la muestra necesario en cada grupo. Para una prueba de la *t* de una muestra, el tamaño de la muestra total es la mitad del número señalado en la lista.

▪ CÁLCULO DE LA VARIABILIDAD

La variabilidad suele comunicarse como la desviación típica o el error típico de la media (ETM). Para calcular el tamaño de la muestra es más útil la desviación típica de la variable. Afortunadamente, es fácil pasar de una medida a otra: la desviación típica es, sencillamente, el error típico multiplicado por la raíz cuadrada de *N*, siendo *N* el número de participantes que componen la media. Suponga un estudio que comunicó que la pérdida de peso de 25 personas que siguieron una dieta con poca fibra fue de 10 ± 2 kg (media ± ETM). La desviación típica sería: $2 \times \sqrt{25} = 10 \text{ kg}$.

▪ FÓRMULA GENERAL PARA OTROS VALORES

A continuación se muestra la fórmula general para otros valores de *E, S,* α y β, o para tamaños de grupo diferentes. Sea:

Z_α = desviación normal típica de α (si la hipótesis alternativa es bilateral, $Z_\alpha = 2{,}58$ cuando $\alpha = 0{,}01$, $Z_\alpha = 1{,}96$ cuando $\alpha = 0{,}05$, y $Z_\alpha = 1{,}645$ cuando $\alpha = 0{,}10$. Si la hipótesis alterna tiva es unilateral, $Z_\alpha = 1{,}645$ cuando $\alpha = 0{,}05$).

Z_β = desviación normal típica de β ($Z_\beta = 0{,}84$ cuando $\beta = 0{,}20$, y $Z_\beta = 1{,}282$ cuando $\beta = 0{,}10$).

q_1 = proporción de participantes en el grupo 1

q_2 = proporción de participantes en el grupo 2

N = número total de participantes necesarios

Entonces:

$N = ([1/q_1 + 1/q_2] \times D^2\ [Z_\alpha + Z_\beta]^2)/E^2$

Los lectores que deseen evitar el trabajo que supone calcular a mano esta fórmula pueden lograr una respuesta instantánea con una calculadora o con nuestra página web (www.epibiostat.ucsf.edu/dcr/). (Como esta fórmula se basa en una aproximación del estadístico *t* con el estadístico *Z*, infravalorará ligeramente el tamaño de la muestra cuando *N* sea menor de 30, aproximadamente. La tabla 6A usa el estadístico *t* para calcular el tamaño de la muestra.)

APÉNDICE 6BTamaño de muestra necesario por grupo cuando se usa la χ^2 o la prueba de la *Z* para comprobar las proporciones de variables dicotómicas
NA

	NÚMERO SUPERIOR:α= 0,05 (UNILATERAL) Oα= 0,10 (BILATERAL);β= 0,20
	NÚMERO MEDIO:α= 0,025 (UNILATERAL) Oα= 0,05 (BILATERAL);β= 0,20
	NÚMERO INFERIOR:α= 0,025 (UNILATERAL) Oα= 0,05 (BILATERAL);β= 0,10
	DIFERENCIA ENTRE P_1 Y P_2

MENOR DE P_1 Y P_2*	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
0,05	381	129	72	47	35	27	22	18	15	13
	473	159	88	59	43	33	26	22	18	16
	620	207	113	75	54	41	33	27	23	19
0,10	578	175	91	58	41	31	24	20	16	14
	724	219	112	72	51	37	29	24	20	17
	958	286	146	92	65	48	37	30	25	21
0,15	751	217	108	67	46	34	26	21	17	15
	944	270	133	82	57	41	32	26	21	18
	1.252	354	174	106	73	53	42	33	26	22
0,20	900	251	121	74	50	36	28	22	18	15
	1.133	313	151	91	62	44	34	27	22	18
	1.504	412	197	118	80	57	44	34	27	23
0,25	1.024	278	132	79	53	38	29	23	18	15
	1.289	348	165	98	66	47	35	28	22	18
	1.714	459	216	127	85	60	46	35	28	23
0,30	1.123	300	141	83	55	39	29	23	18	15
	1.415	376	175	103	68	48	36	28	22	18
	1.883	496	230	134	88	62	47	36	28	23
0,35	1.197	315	146	85	56	39	29	23	18	15

		1.509	395	182	106	69	48	36	28	22	18
		2.009	522	239	138	90	62	47	35	27	22
0,40		1.246	325	149	86	56	39	29	22	17	14
		1.572	407	186	107	69	48	35	27	21	17
		2.093	538	244	139	90	62	46	34	26	21
	0,45	1.271	328	149	85	55	38	28	21	16	13
		1.603	411	186	106	68	47	34	26	20	16
		2.135	543	244	138	88	60	44	33	25	19
0,50		1.271	325	146	83	53	36	26	20	15	—
		1.603	407	182	103	66	44	32	24	18	—
		2.135	538	239	134	85	57	42	30	23	—
0,55		1.246	315	141	79	50	34	24	18	—	—
		1.572	395	175	98	62	41	29	22	—	—
		2.093	522	230	127	80	53	37	27	—	—
0,60		1.197	300	132	74	46	31	22	—	—	—
		1.509	376	165	91	57	37	26	—	—	—
		2.009	496	216	118	73	48	33	—	—	—
0,65		1.123	278	121	67	41	27	—	—	—	—
		1.415	348	151	82	51	33	—	—	—	—
		1.883	459	197	106	65	41	—	—	—	—

0,70	1.024	251	108	58	35	—	—	—	—	—
	1.289	313	133	72	43	—	—	—	—	—
	1.714	412	174	92	54	—	—	—	—	—
0,75	900	217	91	47	—	—	—	—	—	—
	1.133	270	112	59	—	—	—	—	—	—
	1.504	354	146	75	—	—	—	—	—	—
0,80	751	175	72	—	—	—	—	—	—	—
	944	219	88	—	—	—	—	—	—	—
	1.252	286	113	—	—	—	—	—	—	—
0,85	578	129	—	—	—	—	—	—	—	—
	724	159	—	—	—	—	—	—	—	—
	958	207	—	—	—	—	—	—	—	—
0,90	381	—	—	—	—	—	—	—	—	—
	473	—	—	—	—	—	—	—	—	—
	620	—	—	—	—	—	—	—	—	—

Para los cálculos unilaterales se usa el estadístico *Z*.

* P_1 representa la proporción de participantes que se espera que tengan el resultado en un grupo; P_2 en el otro grupo. (En un estudio de casos y testigos, P_1 representa la proporción de casos con la variable predictiva; P_2 la proporción de testigos con la variable predictiva.) Para calcular el tamaño de la muestra, recorra transversalmente desde el menor de P_1 y P_2, y descienda por la *diferencia* esperada entre P_1 y P_2. Los tres números representan el tamaño de la muestra necesario en cada grupo para los valores especificados de α y β.

En la tabla 6B-2 se ofrecen más detalles para P_1 y P_2 entre 0,01 y 0,10.

NÚMERO SUPERIOR:α= 0,05 (UNILATERAL) Oα= 0,10 (BILATERAL);β= 0,20

NÚMERO MEDIO:α= 0,025 (UNILATERAL) Oα= 0,05 (BILATERAL);β= 0,20

NÚMERO INFERIOR:α= 0,025 (UNILATERAL) Oα= 0,05 (BILATERAL);β= 0,10

MENOR DE P_1 Y P_2*	DIFERENCIA ENTRE P_1 Y P_2									
	0,01	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**	**0,10**
0,01	2.019	700	396	271	204	162	134	114	98	87
	2.512	864	487	332	249	197	163	138	120	106
	3.300	1.125	631	428	320	254	209	178	154	135
0,02	3.205	994	526	343	249	193	157	131	113	97
	4.018	1.237	651	423	306	238	192	161	137	120
	5.320	1.625	852	550	397	307	248	207	177	154
0,03	4.367	1.283	653	414	294	224	179	148	126	109
	5.493	1.602	813	512	363	276	220	182	154	133
	7.296	2.114	1.067	671	474	359	286	236	199	172
0,04	5.505	1.564	777	482	337	254	201	165	139	119
	6.935	1.959	969	600	419	314	248	203	170	146
	9.230	2.593	1.277	788	548	410	323	264	221	189
0,05	6.616	1.838	898	549	380	283	222	181	151	129
	8.347	2.308	1.123	686	473	351	275	223	186	159

	11.123	3.061	1.482	902	620	460	360	291	242	206
0,06	7.703	2.107	1.016	615	422	312	243	197	163	139
	9.726	2.650	1.272	769	526	388	301	243	202	171
	12.973	3.518	1.684	1.014	691	508	395	318	263	223
0,07	8.765	2.369	1.131	680	463	340	263	212	175	148
	11.076	2.983	1.419	850	577	423	327	263	217	183
	14.780	3.965	1.880	1.123	760	555	429	343	283	239
0,08	9.803	2.627	1.244	743	502	367	282	227	187	158
	12.393	3.308	1.562	930	627	457	352	282	232	195
	16.546	4.401	2.072	1.229	827	602	463	369	303	255
0,09	10.816	2.877	1.354	804	541	393	302	241	198	167
	13.679	3.626	1.702	1.007	676	491	377	300	246	207
	18.270	4.827	2.259	1.333	893	647	495	393	322	270
0,10	11.804	3.121	1.461	863	578	419	320	255	209	175
	14.933	3.936	1.838	1.083	724	523	401	318	260	218
	19.952	5.242	2.441	1.434	957	690	527	417	341	285

Los cálculos unilaterales usan el estadístico *Z*.

▪ FÓRMULA GENERAL PARA OTROS VALORES

La fórmula general para calcular el tamaño de la muestra total (*N*) necesario para un estudio usando el estadístico *Z*, donde P_1 y P_2 se definen como se señala anteriormente, es el siguiente (en el apéndice 6A se ofrecen la definiciones de Z_α y Z_β). Sea:

q_1 = proporción de participantes del grupo 1

q_2 = proporción de participantes del grupo 2

N = número total de participantes

$P = q_1 P_1 + q_2 P_2$

Los lectores que deseen evitar el trabajo que supone calcular a mano con esta fórmula pueden obtener una respuesta instantánea con una calculadora o en nuestra página web (www.epibiostat.ucsf.edu/dcr/). (Esta fórmula no incluye la corrección de continuidad de Fleiss-Tytun-Ury y, por tanto, infravalora el tamaño de muestra necesario hasta en un 10 %. En las tablas 6B-1 y 6B-2 se incluye esta corrección de continuidad.)

APÉNDICE 6CTamaño total de la muestra necesario cuando se usa el coeficiente de correlación (r)
NA

UNILATERAL	α=	**0,005**			**0,025**			**0,05**		
BILATERAL	α=	**0,01**			**0,05**			**0,10**		
	β=	**0,05**	**0,10**	**0,20**	**0,05**	**0,10**	**0,20**	**0,05**	**0,10**	**0,20**
r*										
0,05		7.118	5.947	4.663	5.193	4.200	3.134	4.325	3.424	2.469
0,10		1.773	1.481	1.162	1.294	1.047	782	1.078	854	616
0,15		783	655	514	572	463	346	477	378	273
0,20		436	365	287	319	259	194	266	211	153
0,25		276	231	182	202	164	123	169	134	98
0,30		189	158	125	139	113	85	116	92	67
0,35		136	114	90	100	82	62	84	67	49
0,40		102	86	68	75	62	47	63	51	37
0,45		79	66	53	58	48	36	49	39	29
0,50		62	52	42	46	38	29	39	31	23
0,60		40	34	27	30	25	19	26	21	16

0,70	27	23	19	20	17	13	17	14	11
0,80	18	15	13	14	12	9	12	10	8

* Para calcular el tamaño total de la muestra, recorra transversalmente desde *r* (el coeficiente de correlación esperado) y descienda desde los valores especificados de α y β.

▪ FÓRMULA GENERAL PARA OTROS VALORES

La fórmula general para otros valores de *r,* α y β es la siguiente (v. en el apéndice 6A las definiciones de Z_α y Z_β . Sea:

r = coeficiente de correlación esperado

$C = 0{,}5 \times \ln [(1 + r)/(1 - r)]$

N = número total de participantes necesarios

Entonces:

$N = [(Z_\alpha + Z_\beta)/C]^2 + 3.$

▪ CÁLCULO DEL TAMAÑO DE LA MUESTRA PARA LA DIFERENCIA ENTRE DOS COEFICIENTES DE CORRELACIÓN

Si se comprueba si un coeficiente de correlación, r_1, es diferente de r_2 (es decir, la hipótesis nula es $r_1 = r_2$; la hipótesis alternativa es $r_1 \neq r_2$), sea:

$C_1 = 0{,}5 \times \ln [(1 + r_1)/(1 - r_1)]$

$C_2 = 0{,}5 \times \ln [(1 + r_2)/(1 - r_2)]$

Entonces:

$N = [(Z_\alpha + Z_\beta/(C_1 - C_2)]^2 + 3.$

APÉNDICE 6DTamaño de la muestra para un estudio descriptivo de una variable continua
NA

	NIVEL DE CONFIANZA		
A/D	**90 %**	**95 %**	**99 %**
0,10	1.083	1.537	2.665
0,15	482	683	1.180
0,20	271	385	664
0,25	174	246	425

0,30	121	171	295
0,35	89	126	217
0,40	68	97	166
0,50	44	62	107
0,60	31	43	74
0,70	23	32	55
0,80	17	25	42
0,90	14	19	33
1,00	11	16	27

[1] *A/D* es la amplitud del intervalo de confianza, calculada como *A* (amplitud total deseada) dividida por *D* (desviación típica de la variable). Para calcular el tamaño total de la muestra, recorra transversalmente desde la *amplitud normalizada* y descienda desde el intervalo de confianza especificado.

▪ FÓRMULA GENERAL PARA OTROS VALORES

Para otros valores de *A* y *D* y un nivel de confianza de (1 - α), el número total de participantes necesarios *(N)* es:

$N = 4Z_{\alpha}^{2}D^{2}/A^{2}$

(V. en el apéndice 6A la definición de Z_{α}.)

APÉNDICE 6ETamaño de la muestra para un estudio descriptivo de una variable dicotómica
NA

NÚMERO SUPERIOR: NIVEL DE CONFIANZA DEL 90 %

NÚMERO MEDIO: NIVEL DE CONFIANZA DEL 95 %

NÚMERO INFERIOR: NIVEL DE CONFIANZA DEL 99 %

AMPLITUD TOTAL DEL INTERVALO DE CONFIANZA (*A*)

PROPORCIÓN ESPERADA (*P*)*	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**0,35**	**0,40**

0,10	98	44	—	—	—	—	—
	138	61	—	—	—	—	—
	239	106	—	—	—	—	—
0,15	139	62	35	22	—	—	—
	196	87	49	31	—	—	—
	339	151	85	54	—	—	—
0,20	174	77	44	28	19	14	—
	246	109	61	39	27	20	—
	426	189	107	68	47	35	—
0,25	204	91	51	33	23	17	13
	288	128	72	46	32	24	18
	499	222	125	80	55	41	31
0,30	229	102	57	37	25	19	14
	323	143	81	52	36	26	20
	559	249	140	89	62	46	35
0,40	261	116	65	42	29	21	16
	369	164	92	59	41	30	23
	639	284	160	102	71	52	40
0,50	272	121	68	44	30	22	17
	384	171	96	61	43	31	24

	666	296	166	107	74	54	42

* Para calcular el tamaño de la muestra, recorra horizontalmente la *proporción esperada (P)* que tiene la variable de interés y descienda desde la *amplitud total (A)* deseada del intervalo de confianza. Los tres números representan el tamaño de la muestra necesario para niveles de confianza del 90 %, 95 % y 99 %.

▪ FÓRMULA GENERAL PARA OTROS VALORES

La fórmula general para otros valores de P y A con un nivel de confianza $(1 - \alpha)$, donde P y A se definen como se ha señalado anteriormente, es la siguiente. Sea:

Z_α = la desviación típica normalizada para una α bilateral, donde $(1 - \alpha)$ es el nivel de confianza (p. ej., ya que $\alpha = 0{,}05$ para un nivel de confianza del 95 %, $Z_\alpha = 1{,}96$; para un nivel de confianza del 90 %, $Z_\alpha = 1{,}65$, y para un nivel de confianza del 99 %, $Z_\alpha = 2{,}58$).

Entonces, el número total de participantes necesario es:

$N = 4Z_\alpha{}^2\mathrm{P}(1 - P)/A^2$

APÉNDICE 6FUso correcto y erróneo de las pruebas de la *t*
NA

Las pruebas de la*t*para dos muestras, el tema principal de este capítulo, se usan para comparar valores medios de una variable en dos grupos de participantes. Los dos grupos pueden definirse mediante una variable predictiva (fármaco activo o placebo en un ensayo clínico aleatorizado, o presencia o ausencia de un factor de riesgo en un estudio de cohortes), o pueden definirse mediante una variable de respuesta, como en un estudio de casos y testigos. Una prueba de la *t* con dos muestras puede ser no emparejada, si se comparan determinaciones obtenidas en una sola ocasión en dos grupos, o emparejada, si se compara el cambio en las determinaciones producido en dos momentos temporales, por ejemplo, antes y después de una intervención, entre dos grupos. Un tercer tipo de prueba de la *t*, la prueba de la*t*pareada con una sola muestra compara el cambio medio de las determinaciones en dos momentos temporales en un solo grupo con cero o con cualquier otro cambio especificado.

La tabla 6F ilustra el uso erróneo de pruebas de la *t* pareadas con una sola muestra en un estudio diseñado para comparaciones entre grupos: un ensayo clínico aleatorizado y con enmascaramiento del efecto de un nuevo hipnótico sobre la calidad de vida. En situaciones como estas, algunos investigadores han realizado (¡y publicado!) hallazgos con dos pruebas de la *t* de una muestra separadas: una en el grupo de tratamiento y una en el del placebo.

En la tabla, los valores de *p* designados con (†) proceden de pruebas de la *t* emparejadas de una muestra. El primer valor de *p* (0,05) muestra un cambio significativo en la calidad de vida en el grupo de tratamiento durante el estudio; el segundo valor de *p* (0,16) no muestra cambios significativos en el grupo testigo. Sin embargo, este análisis no permite realizar inferencias sobre diferencias entre los grupos, y sería erróneo concluir que había un efecto significativo del tratamiento.

Los valores de *p* designados con (*) representan los resultados adecuados de una prueba de la *t* para dos muestras. Los dos primeros valores de *p* (0,87 y 0,64) corresponden a pruebas de la *t* para dos muestras no emparejadas y no muestran diferencias estadísticamente significativas en el grupo en las determinaciones inicial y final de la calidad de vida. El último valor de *p* (0,17) corresponde a una prueba de la *t* para dos muestras emparejadas; está más próximo a 0,05 que el valor de *p* para los valores del fin del estudio (0,64), porque las diferencias de las medias emparejadas tienen menores desviaciones típicas. Sin embargo, la mejora de la calidad de vida en el grupo de tratamiento (1,3) no era significativamente diferente de la del grupo al que se administró el placebo (0,9), y la conclusión correcta es que el estudio no halló que el tratamiento fuera eficaz.

CALIDAD DE VIDA EXPRESADA COMO LA MEDIA ± DT

MOMENTO DE LA DETERMINACIÓN	**TRATAMIENTO (*N* = 100)**	**TESTIGO (*N* = 100)**	**VALOR DE P**
Inicio	7,0 ± 4,5	7,1 ± 4,4	0,87*

Final del estudio	8,3 ± 4,7	8,0 ± 4,6	0,64*
Valor de *p*	0,05†	0,16†	
Diferencia	1,3 ± 2,1	0,9 ±2,0	0,17**

* Comparación entre el grupo de tratamiento y el grupo testigo.

† Comparación entre los valores inicial y final del estudio.

BIBLIOGRAFÍA

1. Lehr R. Sixteen S-squared over D-squared: a relation for crude sample size estimates. Stat Med 1992;11:1099-1102. Citado aquí

2. Barthel FM, Babiker A, Royston P, Parmar MK. Evaluation of sample size and power for multi-arm survival trials allowing for non-uniform accrual, non-proportional hazards, loss to follow-up and cross-over. Stat Med 2006;25(15):2521-2542. Citado aquí

3. Ahnn S, Anderson SJ. Sample size determination in complex clinical trials comparing more than two groups for survival endpoints. Stat Med 1998;17(21):2525-2534. Citado aquí

4. Donner A. Sample size requirements for stratified cluster randomization designs [published erratum appears in Stat Med 1997;30(16):2927]. Stat Med 1992;11:743-750. Citado aquí

5. Kerry SM, Bland JM. Trials which randomize practices II: sample size. Fam Pract 1998;15:84-87. Citado aquí

6. Hemming K, Girling AJ, Sitch AJ, et al. Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. BMC Med Res Methodol 2011;11:102. Citado aquí

7. Jahn-Eimermacher A, Ingel K, Schneider A. Sample size in cluster-randomized trials with time to event as the primary endpoint. Stat Med 2013;32(5):739-751. Citado aquí

8. Edwardes MD. Sample size requirements for case-control study designs. BMC Med Res Methodol 2001;1:11. Citado aquí

9. Drescher K, Timm J, Jöckel KH. The design of case-control studies: the effect of confounding on sample size requirements. Stat Med 1990;9:765-776. Citado aquí

10. Lui KJ. Sample size determination for case-control studies: the influence of the joint distribution of exposure and confounder. Stat Med 1990;9:1485-1493. Citado aquí

11. Latouche A, Porcher R, Chevret S. Sample size formula for proportional hazards modelling of competing risks. Stat Med 2004;23(21):3263-3274. Citado aquí

12. Novikov I, Fund N, Freedman LS. A modified approach to estimating sample size for simple logistic regression with one continuous covariate. Stat Med 2010;29(1):97-107. Citado aquí

13. Vaeth M, Skovlund E. A simple approach to power and sample size calculations in logistic regression and Cox regression models. Stat Med 2004;23(11):1781-1792. Citado aquí

14. Dupont WD, Plummer WD Jr. Power and sample size calculations for studies involving linear regression. Control Clin Trials 1998;19:589-601. Citado aquí

15. Murcray CE, Lewinger JP, Conti DV, et al. Sample size requirements to detect gene-environment interactions in genome-wide

association studies. Genet Epidemiol 2011;35(3):201-210. Citado aquí

16. Wang S, Zhao H. Sample size needed to detect gene-gene interactions using linkage analysis. Ann Hum Genet 2007;71(Pt 6):828-842. Citado aquí

17. Witte JS. Rare genetic variants and treatment response: sample size and analysis issues. Stat Med 2012;31(25): 3041-3050. Citado aquí

18. Willan AR. Sample size determination for cost-effectiveness trials. Pharmacoeconomics 2011;29(11):933-949. Citado aquí

19. Glick HA. Sample size and power for cost-effectiveness analysis (Part 2): the effect of maximum willingness to pay. Pharmacoeconomics 2011;29(4):287-296. Citado aquí

20. Glick HA. Sample size and power for cost-effectiveness analysis (Part 1). Pharmacoeconomics 2011;29(3): 189-198. Citado aquí

21. Patel HI. Sample size for a dose-response study [published erratum appears in J Biopharm Stat 1994;4:127]. J Biopharm Stat 1992;2:l-8. Citado aquí

22. Day SJ, Graham DF. Sample size estimation for comparing two or more treatment groups in clinical trials. Stat Med 1991;10:33-43. Citado aquí

23. Guo JH, Chen HJ, Luh WM. Sample size planning with the cost constraint for testing superiority and equivalence of two independent groups. Br J Math Stat Psychol 2011;64(3):439-461. Citado aquí

24. Zhang P. A simple formula for sample size calculation in equivalence studies. J Biopharm Stat 2003;13(3):529-538. Citado aquí

25. Stucke K, Kieser M. A general approach for sample size calculation for the three-arm 'gold standard' non-inferiority design. Stat Med 2012;31(28):3579-3596. Citado aquí

26. Julious SA, Owen RJ. A comparison of methods for sample size estimation for non-inferiority studies with binary outcomes. Stat Methods Med Res 2011;20(6):595-612. Citado aquí

27. Obuchowski NA. Sample size tables for receiver operating characteristic studies. AJR Am J Roentgenol 2000;175(3):603-608. Citado aquí

28. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol 1991;44:763-770. Citado aquí

29. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85(3):257-268. Citado aquí

30. Jewell NP. Statistics for epidemiology. Boca Raton: Chapman and Hall, 2004, p. 68. Citado aquí

CAPÍTULO 7 Diseño de estudios transversales y de cohortes

Stephen B. Hulley
Steven R. Cummings
Thomas B. Newman

Los estudios observacionales tienen dos finalidades principales: descriptiva, analizar las distribuciones de las variables predictivas y las variables de respuesta en una población, y analítica, caracterizar las asociaciones entre estas variables predictivas y de respuesta. En este capítulo se presentan dos diseños observacionales básicos, que se clasifican por el marco temporal en el que se realizan las mediciones.

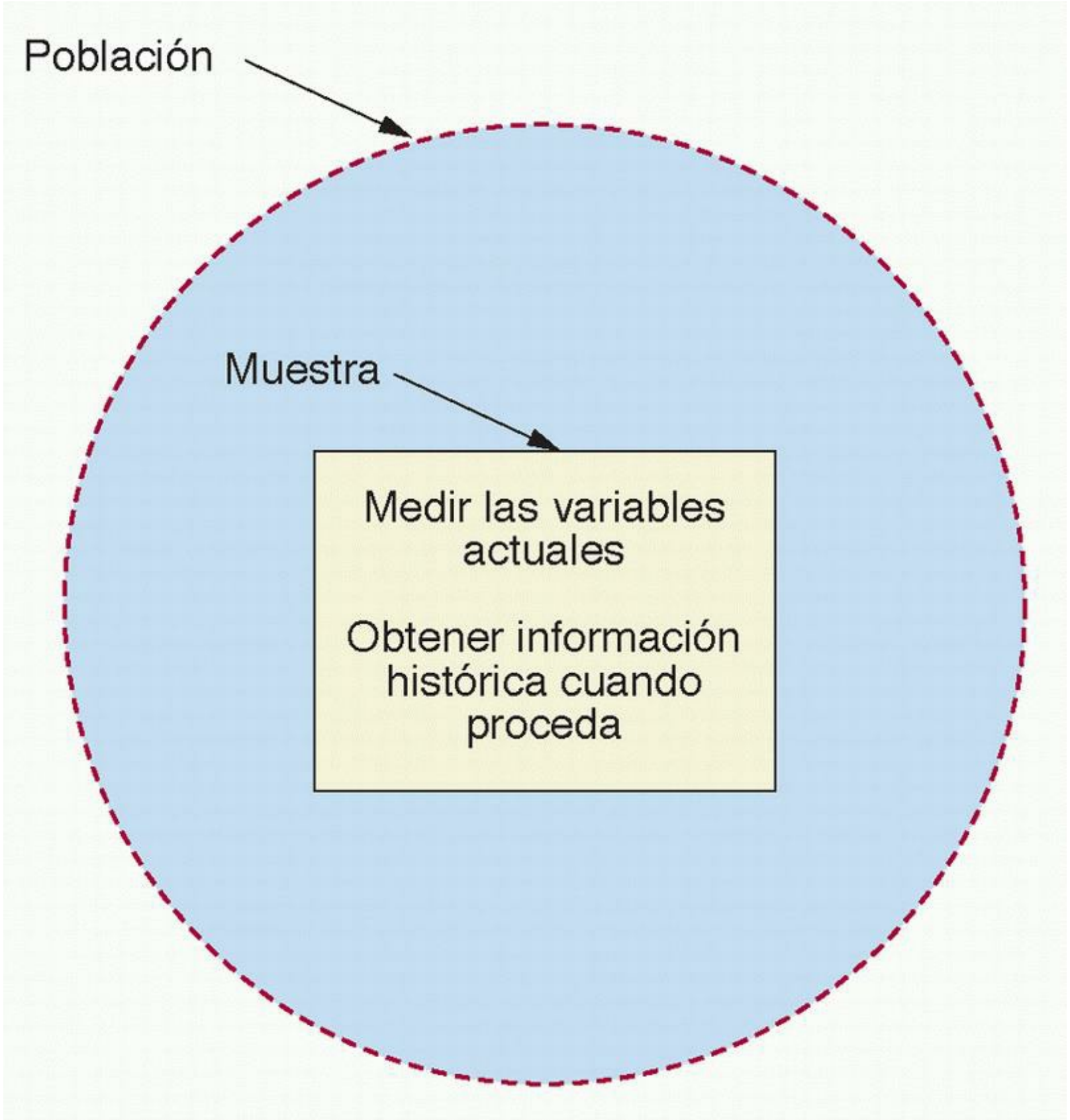
En un estudio transversal, el investigador realiza todas las mediciones en una única ocasión o en un período de tiempo corto. Extrae una muestra de la población y estudia las distribuciones de las variables en esa muestra, en ocasiones denominándolas variables predictivas y variables de resultados, de acuerdo con la credibilidad biológica y la información histórica. Por ejemplo, si está interesado en estudiar la relación entre el peso corporal y la presión arterial, podría medir estas variables en una única visita en la consulta en todos los pacientes del estudio, y explorar si las personas con mayor peso corporal tienen más probabilidad de tener hipertensión.

En un estudio de cohortes, las mediciones se realizan en un período de tiempo en un grupo de participantes a los que se ha identificado al comienzo del estudio («la cohorte»). Por lo tanto, la característica que define los estudios de cohortes es que se sigue longitudinalmente a un grupo reunido al comienzo. Por ejemplo, el investigador podría medir el peso corporal y la presión arterial en una cohorte de personas en estudio en una visita inicial en la consulta y después seguirlas durante 5 años para determinar la relación entre el peso inicial y la incidencia de hipertensión. En este capítulo se analizan los diseños de cohortes prospectivos y retrospectivos, y los diseños de cohortes múltiples. También se abordan los abordajes del análisis estadístico y la importancia de optimizar la retención de la cohorte durante el seguimiento.

▪ ESTUDIOS TRANSVERSALES

En un estudio transversal, todas las mediciones se realizan aproximadamente al mismo tiempo, sin período de seguimiento (fig. 7-1). Los diseños transversales son adecuados para el objetivo de describir variables y sus patrones de distribución. Por ejemplo, en el *National Health and Nutrition Examination Survey* (NHANES), a principios de la década de 1970, se entrevistó y exploró a una muestra diseñada para representar a toda la población estadounidense de 1-74 años. Este estudio transversal fue una importante fuente de información sobre la salud y los hábitos de la población estadounidense el año en que se realizó, y ofreció estimaciones de aspectos como la prevalencia del tabaquismo en diversos grupos demográficos. Posteriormente se han realizado de manera periódica otros estudios NHANES transversales, y todos los conjuntos de datos de los estudios NHANES están disponibles para su uso por el público (www.cdc.gov/nchs/nhanes.htm).

Los estudios transversales se pueden utilizar para explorar asociaciones, aunque la elección de qué variables se van a considerar predictivas y cuáles de respuesta depende de las hipótesis de causa y efecto del investigador, más que del diseño del estudio. Esta elección es fácil para factores constitucionales, como edad, raza y sexo; estos factores no se pueden alterar por otras variables, por lo que siempre son factores predictivos. Sin embargo, para otras variables, la elección puede ir en ambos sentidos. Por ejemplo, en el estudio NHANES III hubo una asociación transversal entre la obesidad infantil y las horas que se pasaba viendo la televisión (1). El que se considere que la obesidad o el tiempo viendo la televisión sea la variable predictiva y la otra la variable de respuesta depende de la hipótesis causal del investigador.



▪FIGURA 7.1. En un estudio transversal, los pasos son:
• Definir los criterios de selección e incluir una muestra de la población.
• Medir los valores actuales de las variables predictivas y de respuesta, muchas veces complementados con información histórica.
Herramientas de imágenes

Al contrario de los estudios de cohortes, que tienen una dimensión temporal longitudinal que se puede utilizar para estimar la incidencia (la proporción que *llega a presentar* una enfermedad a lo largo del tiempo), los estudios transversales ofrecen información sobre la prevalencia, la proporción que *tiene* una enfermedad o trastorno en un momento determinado. La prevalencia le importa al médico, que debe estimar la probabilidad de que el paciente que está en su consulta tenga una enfermedad concreta; cuanto mayor sea la prevalencia, mayor será la «probabilidad previa» de la enfermedad (la probabilidad antes de que se disponga de los resultados de las diversas pruebas diagnósticas; cap. 12). Es este el motivo por el que más pacientes con dolor de rodilla tienen artrosis que reumatismo palindrómico. La

prevalencia también es útil para los planificadores sanitarios, que quieren saber cuántas personas tienen determinadas enfermedades para poder asignar suficientes recursos para atenderlas. Cuando se analizan estudios transversales, se puede comparar la prevalencia del resultado en los que tienen y no tienen una exposición, lo que permite obtener la prevalencia relativa del resultado, el equivalente transversal del riesgo relativo (pueden verse ejemplos en el apéndice 8A).

En ocasiones, los estudios transversales describen la prevalencia de haber realizado algo o de haber tenido alguna vez una enfermedad o trastorno. En este caso, es importante asegurarse de que el tiempo de seguimiento sea igual en las personas expuestas y no expuestas. Esto se ilustra en el ejemplo 7-1 , en el que se analizó la prevalencia de haber probado alguna vez el tabaco en un estudio transversal de niños con diferentes niveles de exposición a películas en las que los actores fuman. Por supuesto, los niños que habían visto más películas también eran mayores, por lo que habían tenido más tiempo para probar el tabaco, por lo que era importante ajustar la edad en los análisis multivariados (cap. 9).

Puntos fuertes y débiles de los estudios transversales

Una importante ventaja de los estudios transversales es que no hay que esperar a que se produzca el resultado. Esto hace que sean rápidos y económicos, y evita el problema de las pérdidas durante el seguimiento. Otra ventaja es que un estudio transversal se puede incluir como primer paso en un estudio de cohortes o un estudio clínico, con un coste añadido escaso o nulo. Los resultados definen las características demográficas y clínicas del grupo en estudio en situación inicial, y en ocasiones pueden mostrar asociaciones de interés transversales.

Sin embargo, como ya se ha señalado, muchas veces es difícil establecer relaciones causales a partir de datos transversales. Los estudios transversales tampoco son prácticos para el estudio de enfermedades poco frecuentes, salvo que la muestra se extraiga de una población de pacientes enfermos y no de la población general. Una serie de casos de este tipo es más adecuada para describir las características de la enfermedad que para analizar diferencias entre estos pacientes y las personas sanas, aunque las comparaciones informales con la experiencia previa en ocasiones permiten identificar factores de riesgo muy potentes. Por ejemplo, en una serie de casos de los 1 000 primeros pacientes con sida, 727 eran hombres homosexuales o bisexuales y 236 consumían drogas por vía intravenosa (3). No hizo falta un grupo testigo formal para concluir que estos grupos tenían aumento del riesgo. Además, en una muestra de personas con una enfermedad puede haber asociaciones de interés, como el mayor riesgo de sarcoma de Kaposi en pacientes con sida que eran homosexuales que en los que consumían drogas inyectables.

EJEMPLO 7-1. Estudio transversal

Sargent y cols. (2) intentaron determinar si la exposición a películas en las que los actores fuman se asocia al inicio del tabaquismo. Los pasos en la realización del estudio fueron:

1. Definir los criterios de selección y reunir la muestra de la población. Los investigadores realizaron un estudio con un sistema de marcación telefónica aleatoria de 6 522 niños estadounidenses de 10 a 14 años de edad.

2. Medir las variables predictiva y de respuesta. Cuantificaron el tabaquismo en 532 películas populares, y a cada uno de los participantes le preguntaron cuáles de un subgrupo seleccionado aleatoriamente de 50 películas habían visto. También se preguntó a los participantes por diversas covariables, como edad, raza, sexo, tabaquismo y nivel educativo de los padres, búsqueda de sensaciones (p. ej., «Me gusta hacer cosas peligrosas») y autoestima (p. ej., «Ojalá fuera otra persona»). La variable de respuesta fue si el niño había intentado alguna vez fumar un cigarrillo.

La prevalencia de haber intentado alguna vez fumar varió desde el 2 % en el cuartil inferior de exposición al tabaco en las películas hasta el 22 % en el cuartil superior. Después de ajustar la edad y otros factores de confusión, estas diferencias eran estadísticamente significativas; los autores estimaron que el 32 % del inicio del tabaquismo se podía atribuir a la exposición a películas en las que los actores fuman.

Como los estudios transversales miden únicamente la prevalencia y no la incidencia, es importante tener precaución cuando se extraigan inferencias sobre las causas, el pronóstico o la evolución natural de una enfermedad. Un factor que se asocia a la prevalencia de una enfermedad puede ser una causa de la enfermedad, aunque también se podría asociar simplemente a la duración de la enfermedad. Por ejemplo, la prevalencia de la insuficiencia renal crónica depende no solo de su incidencia, sino también de su supervivencia una vez que se ha producido. A la vista de la observación de que la obesidad se asocia a mayor supervivencia de pacientes en diálisis (4), en un estudio transversal de los factores predictivos de la insuficiencia renal crónica se podría sobrestimar la asociación entre obesidad e insuficiencia renal.

Estudios en serie

En ocasiones, los investigadores realizan una serie de estudios transversales en la misma población, por ejemplo, cada 5 años. Se puede

utilizar este diseño para extraer inferencias sobre los patrones cambiantes a lo largo del tiempo. Por ejemplo, Zito y cols. (5), utilizando estudios transversales anuales, describieron que la prevalencia del consumo de fármacos psicótropos de venta con receta en jóvenes (< 20 años de edad) había aumentado más de tres veces entre 1987 y 1996 en una población de la región del Atlántico medio atendida por Medicaid. Los estudios transversales en serie tienen un marco temporal longitudinal, pero no son lo mismo que un estudio de cohortes, porque cada vez se extrae una nueva muestra. En consecuencia, no se pueden evaluar los cambios que se producen en las personas, y los hallazgos se pueden ver modificados por las personas que entran o salen de la población (y, por lo tanto, de las muestras) por nacimientos, muertes y movimientos migratorios.

Volver al principio

▪ ESTUDIOS DE COHORTES

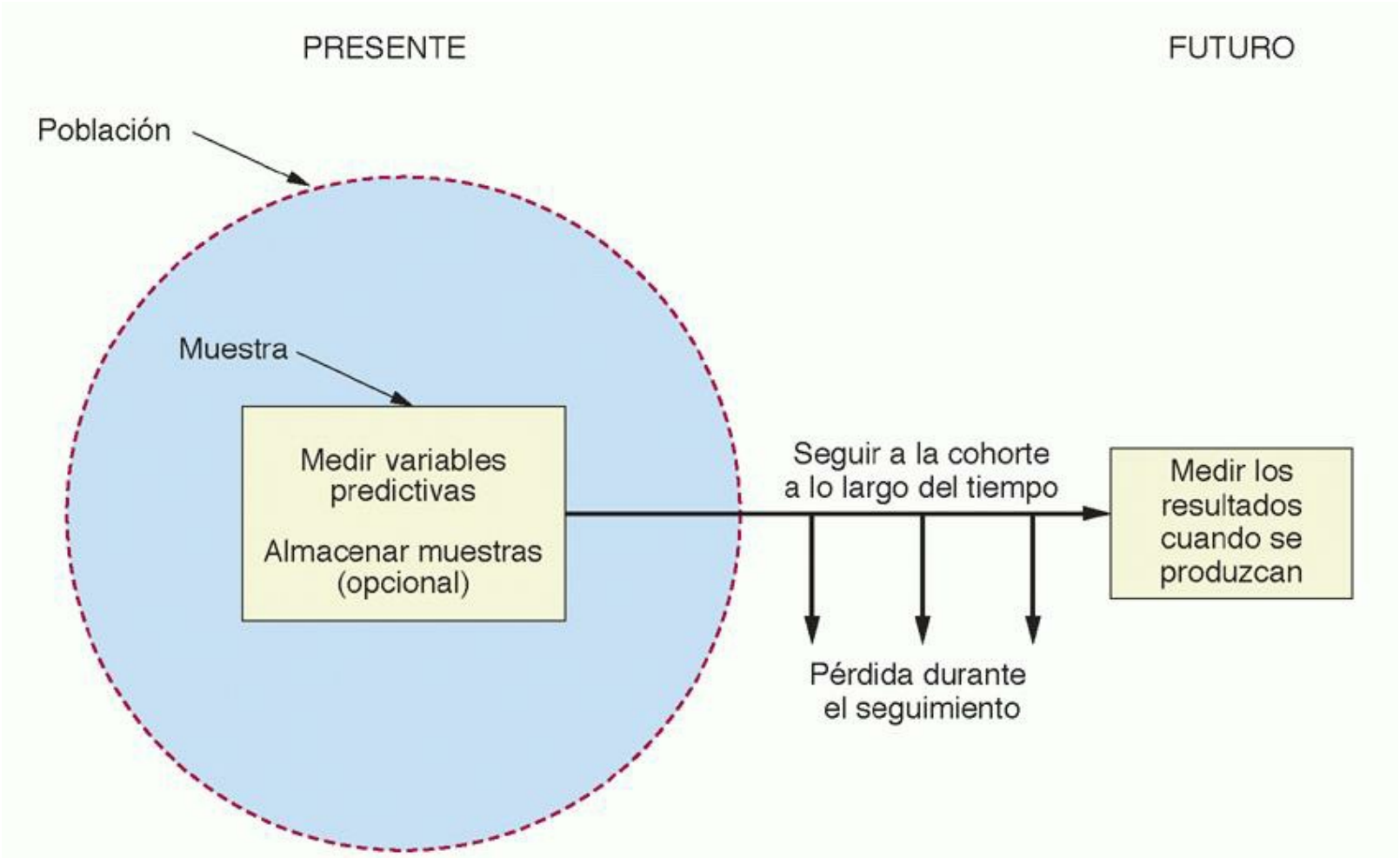
Estudios de cohortes prospectivos

Cohorte era el término romano que designaba a un grupo de soldados que avanzaban juntos, y en las investigaciones clínicas una cohorte es un grupo de personas, especificado al comienzo del estudio y al que se sigue a lo largo del tiempo. En un estudio de cohortes prospectivo, el investigador empieza reuniendo una muestra de participantes (fig. 7-2). Mide en cada participante características que podrían predecir los resultados subsiguientes, y sigue a estas personas, realizando determinaciones periódicas de las variables de respuesta de interés (ejemplo 7-2).

Puntos fuertes y puntos débiles de los estudios de cohortes prospectivos

Una importante ventaja del diseño de cohortes es que, al contrario de los diseños transversales, permite el cálculo de la incidencia, el número de nuevos casos de una enfermedad que se producen a lo largo del tiempo (tabla 7-1). La medición de los niveles del factor predictivo antes de que se produzca el resultado establece la secuencia temporal de las variables, lo que refuerza el proceso de inferir la base causal de una asociación. El enfoque prospectivo también evita que las variables predictivas se vean modificadas por el resultado o por el conocimiento de su aparición, y permite que el investigador mida variables de forma más completa y exacta de lo que es posible si se hace retrospectivamente. Esto es importante para variables predictivas como los hábitos dietéticos, que es difícil que los participantes recuerden con exactitud. Cuando se estudian retrospectivamente enfermedades mortales, las mediciones de las variables predictivas en el fallecido solo pueden reconstruirse a partir de fuentes indirectas, como las historias clínicas o datos de amigos y familiares.

Todos los estudios de cohortes comparten el inconveniente general de los estudios de observación (en comparación con los ensayos clínicos) de que la inferencia causal es un reto y la interpretación a menudo se enturbia por las influencias de las variables de confusión (cap. 9). Un punto débil concreto del diseño prospectivo es su coste y su ineficacia para estudiar variables de respuesta poco frecuentes. Incluso enfermedades en las que pensamos como relativamente frecuentes, como el cáncer de mama, aparecen con una incidencia tan baja en cualquier año determinado que debe seguirse a una gran cantidad de personas durante largos períodos de tiempo para observar suficientes respuestas a fin de obtener resultados significativos. Los diseños de cohortes son más eficaces para variables de respuesta dicotómicas, que son más frecuentes e inmediatas, y para variables de respuesta continuas.



▪FIGURA 7-2. En un estudio de cohortes prospectivo, los pasos son:
• Definir los criterios de selección e incluir una muestra de la población («la cohorte»).
• Medir las variables predictivas y, si procede, el nivel inicial de la variable de respuesta.
• Plantear la opción de almacenar muestras, imágenes, etc., para el análisis posterior de las variables predictivas.
• Seguir a la cohorte a lo largo del tiempo, minimizando las pérdidas durante el seguimiento.
• Medir las variables de respuesta durante el seguimiento.
Herramientas de imágenes

EJEMPLO 7-2. Estudio de cohortes prospectivo

En el clásico estudio *Nurse's Health Study* se examinó la incidencia y los factores de riesgo de enfermedades habituales en las mujeres. Los pasos básicos para realizar este estudio fueron:

1. Definir los criterios de selección y reunir la cohorte. En 1976, los investigadores consiguieron listas de enfermeras colegiadas de 25 a 42 años de edad en los 11 estados más poblados y les enviaron una invitación para participar en el estudio; las que aceptaron constituyeron la cohorte.

2. Medir las variables predictivas, incluyendo los posibles factores de confusión. Enviaron un cuestionario sobre peso, ejercicio y otros posibles factores de riesgo, y recibieron cuestionarios cumplimentados por 121 700 enfermeras. Enviaron cuestionarios periódicamente para preguntar sobre factores de riesgo adicionales y actualizar el estado de algunos que se habían determinado con anterioridad.

3. Realizar un seguimiento de la cohorte y medir las respuestas. Los cuestionarios periódicos incluían también preguntas sobre la incidencia de diversas respuestas de enfermedad, que fueron validadas por los investigadores.

El enfoque prospectivo permitió que los investigadores realizaran determinaciones al principio y recopilaran datos sobre variables de respuesta posteriores. El gran tamaño de la cohorte y el prolongado período de seguimiento han proporcionado una potencia estadística elevada para estudiar factores de riesgo de sufrir cáncer y otras enfermedades.

Por ejemplo, los investigadores examinaron la hipótesis de que el aumento de peso incrementa el riesgo de que una mujer sufra cáncer de mama después de la menopausia (6). Las mujeres notificaron su peso a los 18 años en un cuestionario inicial y el peso durante el seguimiento en los cuestionarios posteriores. Los investigadores lograron realizar el seguimiento del 95 % de las mujeres, y se

confirmaron 1 517 casos de cáncer de mama durante los siguientes 12 años. Las mujeres de mayor peso tuvieron mayor riesgo de sufrir cáncer de mama tras la menopausia, y las que aumentaron más de 20 kg desde los 18 años de edad presentaban un riesgo doble de sufrir cáncer de mama (riesgo relativo = 2; intervalo de confianza al 95 %, 1,4 a 2,8). El ajuste realizado por posibles factores de confusión no modificó el resultado.

Estudios de cohortes retrospectivos

El diseño de un estudio de cohortes retrospectivo (fig. 7-3) difiere del de uno prospectivo en que la reunión de la cohorte, las determinaciones basales y el seguimiento de control ya se han producido en el pasado. Este tipo de estudio solo es posible si se dispone de datos adecuados sobre las variables predictivas en una cohorte de personas a las que se ha reunido por otros motivos, como una base de datos clínica o administrativa electrónica (ejemplo 7-3).

TIPO DE ESTUDIO	ESTADÍSTICO	DEFINICIÓN
Transversal	Prevalencia	$\frac{\text{Número de personas que tienen una enfermedad o trastorno en un momento determinado}}{\text{Número de personas en riesgo}}$
Cohortes	Tasa de incidencia	$\frac{\text{Número de personas que contraen una enfermedad o trastorno}}{\text{Número de personas en riesgo} \times \text{período temporal en riesgo}}$

Puntos fuertes y puntos débiles de los estudios de cohortes retrospectivos

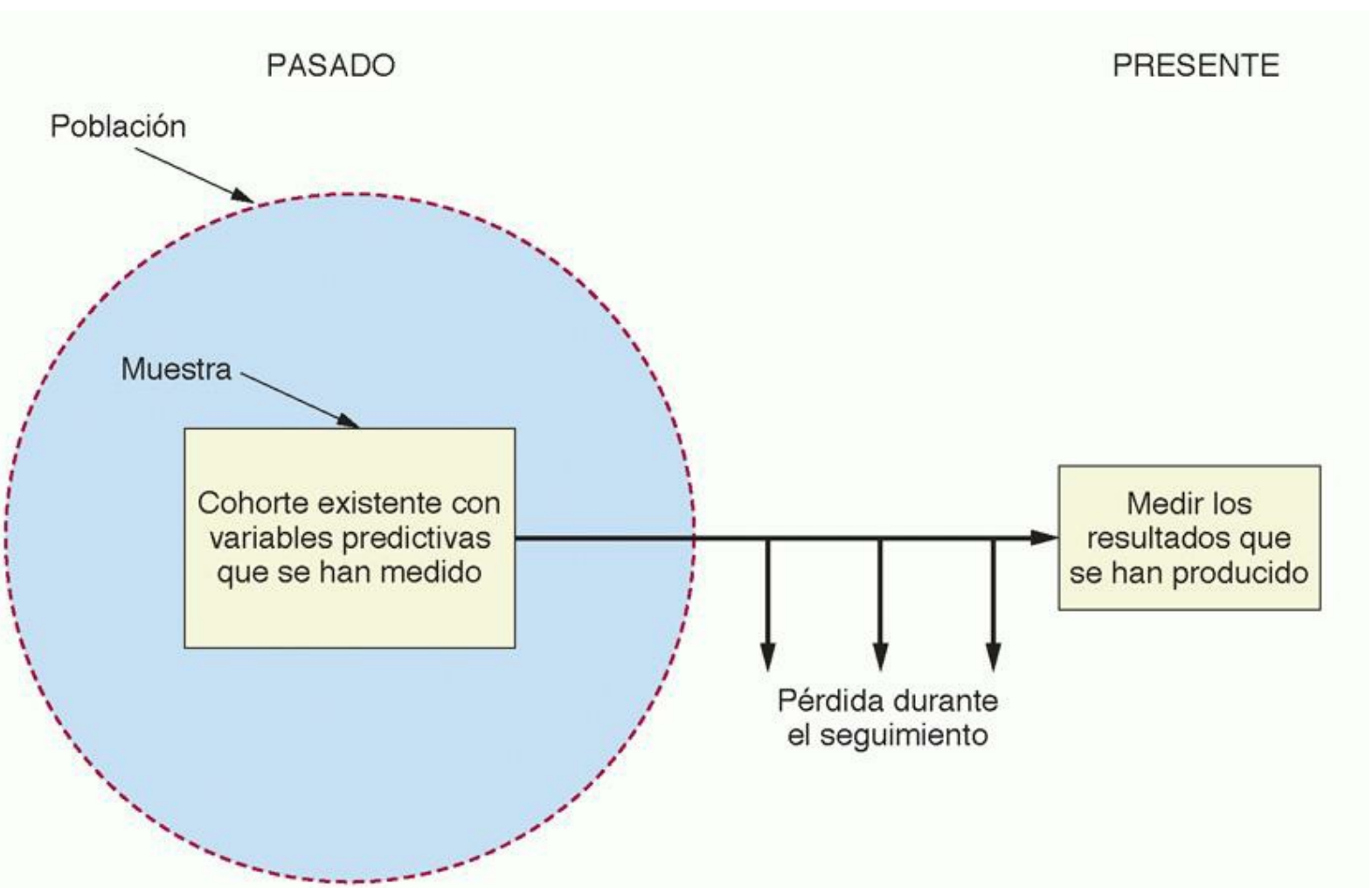
Los estudios de cohortes retrospectivos tienen muchos de los puntos fuertes de los estudios de cohortes prospectivos, y tienen la ventaja de ser mucho menos costosos y requerir menos tiempo. Los participantes ya están reunidos, las mediciones basales ya se han realizado y el período de seguimiento ya ha tenido lugar. Los principales inconvenientes son el escaso control que el investigador tiene sobre el método de muestreo y el seguimiento de la población, y sobre la naturaleza y la calidad de las mediciones iniciales. Los datos existentes pueden ser incompletos o inexactos, o haberse medido de formas no ideales para responder a la pregunta de la investigación.

Estudios de cohortes múltiples y testigos externos

Los estudios de cohortes múltiples comienzan con dos o más muestras separadas de participantes, habitualmente un grupo con exposición a un posible factor de riesgo y uno o más grupos sin exposición o con un nivel de exposición menor (fig. 7-4). Después de definir cohortes adecuadas con diferentes niveles de exposición a la variable predictiva de interés, el investigador mide otras variables predictivas, sigue a las cohortes y evalúa los resultados, como en cualquier otro tipo de estudio de cohortes (ejemplo 7-4).

El uso de dos muestras diferentes de participantes en un diseño de cohorte doble no se debe confundir con el uso de dos muestras en el diseño de casos y testigos (cap. 8). En un estudio de cohorte doble se eligen los dos grupos de participantes por el nivel de un factor predictivo, mientras que en un estudio de casos y testigos se eligen los dos grupos basándose en la presencia o la ausencia de un resultado.

En una variación del diseño de cohortes múltiples, se puede comparar la frecuencia de un resultado en una cohorte con las frecuencias del resultado en datos de censos o de registros procedentes de diferentes poblaciones. Por ejemplo, en un estudio clásico de si los mineros del uranio tenían mayor incidencia de cáncer de pulmón, Wagoner y cols. (10) compararon la incidencia de cánceres respiratorios en 3 415 mineros del uranio con la de personas blancas que vivían en los mismos estados. La mayor incidencia de cáncer de pulmón que se observó en los mineros ayudó a establecer que la exposición laboral a las radiaciones ionizantes es una importante causa de cáncer de pulmón.



▪FIGURA 7-3. En un estudio de cohortes retrospectivo, la selección de la cohorte y el seguimiento se han realizado en el pasado, por lo que los pasos son:
• Identificar una cohorte existente que tenga alguna información predictiva que ya se haya registrado.
• Evaluar las pérdidas durante el seguimiento que se han producido.
• Medir las variables de respuesta que ya se han producido.
Herramientas de imágenes

EJEMPLO 7-3. Estudio de cohortes retrospectivo

Pearce y cols. utilizaron datos del *UK National Health Service Central Registry* para describir el riesgo de leucemia y tumores cerebrales asociados a la realización de TC craneal durante la infancia (7). Los pasos en la realización del estudio fueron:

1. Identificar una cohorte existente adecuada. La cohorte estuvo formada por 178 644 niños y adultos jóvenes de < 2 años de edad a los que se habían realizado TC craneales entre 1985 y 2002.

2. Recoger datos de la variable predictiva. Los investigadores revisaron las historias para recoger el sexo, la edad, el número y el tipo de técnicas radiológicas y la dosis de radiación estimada.

3. Recoger datos de la variable de respuesta. Para evitar la inclusión de TC relacionadas con un diagnóstico de cáncer, los investigadores registraron la leucemia que se produjo al menos 2 años después de la primera TC, y los tumores cerebrales al menos 5 años después de la primera TC, hasta 2008.

Las TC durante la infancia aumentaron significativamente el riesgo de leucemia y cáncer cerebral, y la incidencia se relacionó con la dosis. Dosis acumuladas de 50-60 mGy aumentaron al triple el riesgo de leucemia y cáncer cerebral. Sin embargo, el aumento absoluto del riesgo fue bajo, un caso de más de cada una de las respuestas por cada 10 000 estudios craneales. Aunque los autores señalaron que los beneficios de las TC probablemente superaron a estos riesgos, también insistieron en que se utilicen las menores dosis de radiación por TC en niños, y que se deben plantear, cuando sea posible, procedimientos alternativos que eviten las radiaciones ionizantes.

Puntos fuertes y puntos débiles de los diseños de cohortes múltiples

El diseño de cohortes múltiples puede ser el único abordaje viable para estudiar exposiciones infrecuentes a posibles riesgos laborales y

ambientales. La utilización de datos de un censo o registro como grupo testigo externo tiene la ventaja adicional de basarse en la población y ser económico. Por lo demás, los puntos fuertes de este diseño son similares a los de otros estudios de cohortes.

EJEMPLO 7-4. Diseño de cohortes múltiples

Para determinar si la ictericia neonatal y la deshidratación significativas tienen efectos adversos sobre el desarrollo neurológico, los investigadores del hospital UCSF and Kaiser Permanente of Northern California (8 , 9) emprendieron un estudio con cohortes triples. Los pasos en la realización del estudio fueron:

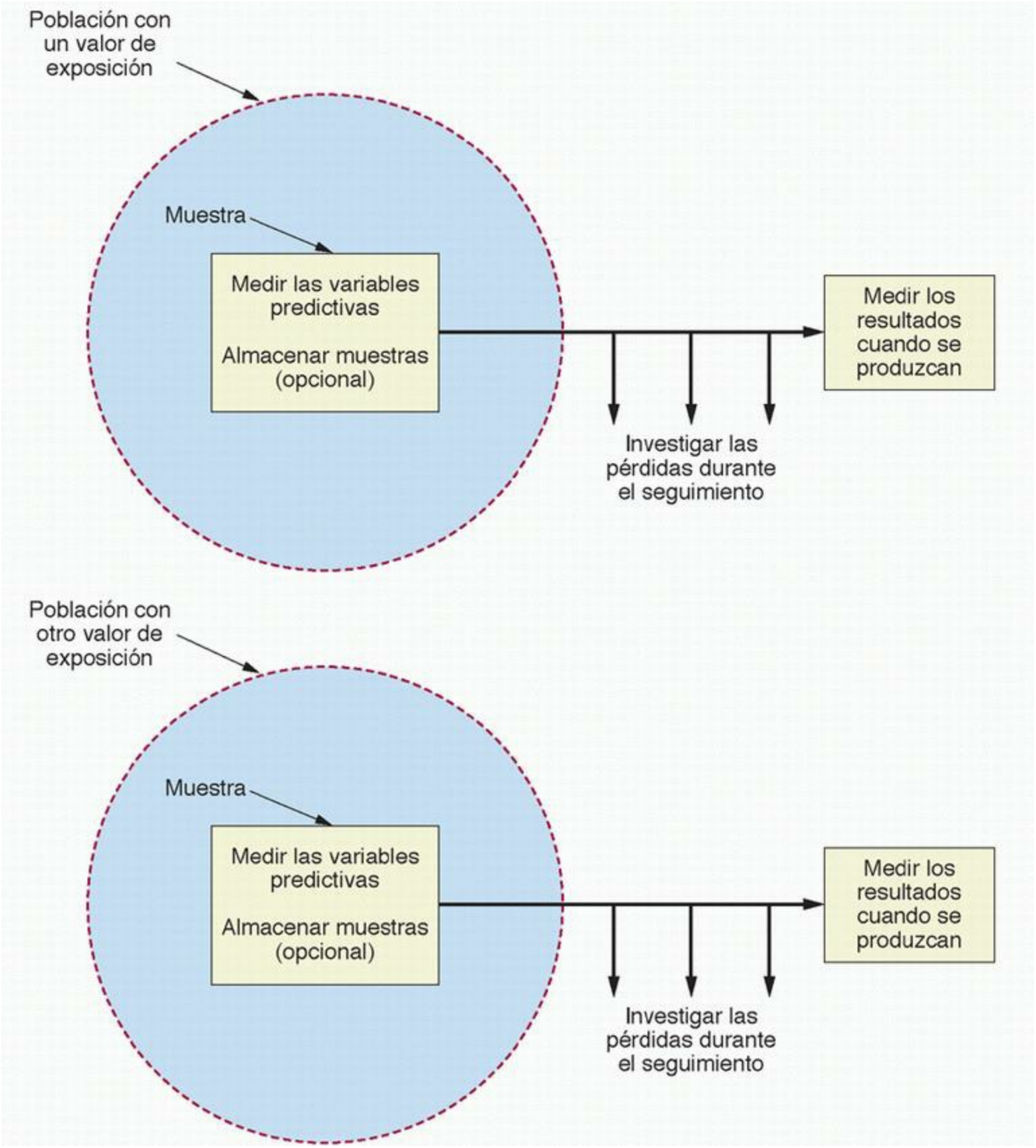
1. Identificar cohortes con diferentes exposiciones. Los investigadores usaron bases de datos electrónicas para identificar recién nacidos a término y casi a término que 1. presentaron una concentración máxima de bilirrubina sérica total ≥ 25 mg/dl, o

2. reingresaron por deshidratación con un sodio sérico ≥ 150 mEq/l o pérdida de peso ≥ 12 % desde el nacimiento, o

3. fueron seleccionados aleatoriamente de la cohorte del nacimiento.

2. Recoger datos de resultados. Los investigadores usaron bases de datos electrónicas para buscar diagnósticos de trastornos neurológicos y realizaron exploraciones neurológicas completas a la edad de 5 años en pacientes que habían dado su consentimiento (con enmascaramiento sobre a cuál de las tres cohortes pertenecía el participante).

Ni la hiperbilirrubinemia ni la deshidratación se asociaron a respuestas adversas.



▪FIGURA 7-4. En un estudio de cohortes dobles (que se puede realizar prospectiva o retrospectivamente), los pasos son:
• Seleccionar dos o más cohortes de poblaciones con diferentes niveles de la exposición (variable predictiva principal).
• Medir otras variables predictivas.
Medir las variables de respuesta durante el seguimiento.
Herramientas de imágenes

El problema de la confusión se acentúa en los estudios de cohortes múltiples, porque las cohortes se reúnen a partir de poblaciones distintas que pueden diferir en aspectos importantes (aparte de la exposición a la variable predictiva) que pueden influir en los resultados. Aunque algunas de estas diferencias, como la edad y la raza, se pueden emparejar o utilizar para ajustar estadísticamente los hallazgos, otras características pueden no ser medibles y crean problemas en la interpretación de las asociaciones observadas.

Volver al principio

▪ ABORDAJE ESTADÍSTICO DE LOS ESTUDIOS DE COHORTES

Los riesgos, las oportunidades y las tasas de incidencia son estimaciones de la frecuencia de un resultado dicotómico en personas a las que se ha seguido durante un período de tiempo. Estas tres mediciones están estrechamente relacionadas, de manera que comparten el mismo numerador, el número de personas que llegan a presentar el resultado dicotómico. Está implícito en estas tres medidas el concepto de estar *en riesgo*, lo que significa que la persona no tenía todavía el resultado de interés al comienzo del estudio. En un estudio prospectivo de los factores predictivos de la diabetes, una mujer que tuviera diabetes en situación inicial no tendría riesgo, porque ya tendría el resultado de interés. Por otro lado, hay enfermedades episódicas, como la insuficiencia cardíaca que precisa ingreso hospitalario, en las que el resultado de interés puede ser la aparición «incidente» de un nuevo episodio, aunque se produzca en alguien que ya tiene la enfermedad.

Considérese un estudio de 1 000 personas a las que se siguió durante 2 años para ver quién presentaba cáncer de pulmón, y en las cuales se produjeron ocho nuevos casos cada año. En la tabla 7-2 se muestra el riesgo, la oportunidad y la incidencia.

De las tres medidas, el riesgo es el más fácil de comprender, debido a su familiaridad cotidiana: el riesgo de tener cáncer de pulmón en 2 años era de 16 de cada 1 000. Las oportunidades son más difíciles de comprender intuitivamente: la oportunidad de contraer cáncer de pulmón era de 16 a 984; afortunadamente, para resultados infrecuentes (como en este caso), las oportunidades son cuantitativamente similares al riesgo y no tienen ninguna ventaja particular. En estudios en los que se comparan dos grupos, la razón de posibilidades***(odds ratio)*** también es similar al cociente de riesgo*cuando la variable de respuesta es poco frecuente*, y este hecho tiene una importancia específica en dos situaciones: forma la base de los cálculos de regresión logística y se utiliza para obtener una aproximación del riesgo relativo en estudios de casos y testigos (apéndice 8B). Las tasas de incidencia, que tienen en consideración la acumulación de episodios a lo largo del tiempo, se expresan como números de episodios divididos por el número de personas-tiempo en riesgo, la cantidad total de seguimiento para cada uno de los participantes en el estudio mientras esa persona esté viva, permanezca en el estudio y no haya tenido todavía el resultado.

En algunos estudios de cohortes puede haber una elevada frecuencia de casos de pérdida durante el seguimiento, seguimiento desigual o muerte, u otros acontecimientos que impiden la determinación de la variable de respuesta. En estos casos es útil comparar las tasas de incidencia entre los grupos: el número de resultados dividido por el número de personas-tiempo en riesgo. Cada participante en el estudio contribuye con un número determinado de meses o años de personas-tiempo desde su inclusión en la cohorte hasta que presenta el resultado de interés o se le «censura» por pérdida durante el seguimiento o muerte. La tasa de incidencia en cualquier grupo del estudio es el número de resultados en ese grupo dividido por la suma de las personas-tiempo en riesgo de ese grupo. Como ocurre para el cociente de riesgo (también conocido como riesgo relativo), el cociente de incidencias se puede estimar como el cociente de las incidencias en personas que tienen y que no tienen un factor de riesgo particular. El modelo de riesgos proporcionales de Cox ofrece un método para el análisis multivariado de los datos de este tipo (en ocasiones denominados datos de «tiempo hasta un acontecimiento»); permite la estimación de cocientes de riesgos instantáneos, que son similares a los cocientes de incidencia y han llegado a utilizarse de manera generalizada como medida de la asociación en los análisis de regresión de Cox.

ESTADÍSTICO	FÓRMULA		EJEMPLO
Riesgo	*N que llegan a presentar el resultado* / *N en riesgo*	16 / 1 000	= 0,016
Oportunidad	*N que llegan a presentar el resultado* / *N que no presentan el resultado*	16 / 984	= 0,0 163
Tasa de incidencia[1]	*N que llegan a presentar el resultado*	16 casos	= 0,008 casos/persona-año

	Personas-tiempo en riesgo	1 992 personas-año	

[1] El denominador de la incidencia es el número en riesgo en el primer año (1 000), más el número en riesgo el segundo año (992).

Otros aspectos sobre los estudios de cohortes

La característica fundamental de un estudio de cohortes es la necesidad de definir la cohorte de personas al *principio* de un período de seguimiento. Los participantes deben ser adecuados para la pregunta de la investigación y deben estar disponibles para poder realizar el seguimiento. Deben parecerse lo suficiente a la población a la que se generalizarán los resultados. El número de participantes debe proporcionar una potencia adecuada.

La calidad del estudio dependerá de la precisión y la exactitud de las determinaciones de las variables predictivas y de respuesta (cap. 4). La capacidad para obtener inferencias sobre causa y efecto también dependerá del grado en que el investigador haya medido todos los posibles factores de confusión (cap. 9), y la capacidad de generalizar a subgrupos de la población dependerá del grado en el que el investigador haya medido todas las fuentes de modificación del efecto. Las variables predictivas pueden cambiar durante el estudio; el que deban o no repetirse las determinaciones, y con qué frecuencia debe hacerse, dependerá del coste, de cuánto es probable que cambie la variable y de la importancia para la pregunta de la investigación de la observación de estos cambios. Las respuestas deben evaluarse mediante criterios normalizados y, cuando su evaluación pueda depender del conocimiento de factores de riesgo fundamentales, es útil que las personas que realicen las evaluaciones desconozcan esa variable predictiva.

Es importante el seguimiento de toda la cohorte, y los estudios prospectivos deben dar una serie de pasos para lograr ese objetivo (tabla 7-3). Se debe excluir desde el comienzo a los participantes que tengan previsto desplazarse fuera del ámbito del estudio o a los que vaya a ser difícil seguir por otros motivos. El investigador debe conseguir pronto información por la cual pueda encontrar a los participantes que cambien de domicilio o que fallezcan, incluyendo la dirección, el número de teléfono y la dirección de correo electrónico del participante, su médico personal y al menos dos amigos o familiares próximos que no vivan en el mismo domicilio. Los números de teléfono móvil y las direcciones de correo electrónico personales son particularmente útiles, porque habitualmente no cambian cuando los participantes, sus amigos o familiares se mudan de domicilio o cambian de trabajo. Cuando sea posible, obtener el número de seguridad social ayudará a determinar la situación vital de los participantes a los que se pierda durante el seguimiento, igual que obtener información de altas hospitalarias de la administración de la seguridad social de los pacientes que reciban atención a través de Medicare. Un contacto periódico con el participante una o dos veces al año ayuda a seguir su trayectoria, y puede aumentar la oportunidad y la exactitud para registrar las respuestas de interés. Para encontrar a los participantes para las evaluaciones de seguimiento a veces se deben realizar esfuerzos persistentes y repetidos por correo postal, correo electrónico, teléfono o incluso visitas domiciliarias.

Volver al principio

▪ RESUMEN

- En un estudio transversal, todas las variables se miden en un único punto temporal, sin distinción estructural entre variables predictivas y variables de respuesta. Los estudios transversales ofrecen datos de causalidad más débiles que los estudios de cohortes, porque no se puede demostrar que la variable predictiva preceda a la variable de respuesta.
- Los estudios transversales son útiles para obtener información descriptiva sobre la prevalencia, y tienen la ventaja de evitar el tiempo, el gasto y los problemas de abandono de un diseño de seguimiento; muchas veces son útiles como primer paso de un estudio de cohortes o un estudio experimental, y se los puede vincular a estudios en serie con muestras independientes para mostrar cambios poblacionales a lo largo del tiempo.
- Los estudios transversales precisan un gran tamaño de la muestra cuando se estudian enfermedades y variables poco frecuentes en la población general, aunque pueden ser útiles en una serie de casos de una enfermedad poco frecuente.

Durante la inclusión

---

1. Excluir aquellos que es probable que se pierdan:

1. Plan de mudanza
2. Duda sobre la voluntad de regresar
3. Mala salud o enfermedad mortal no relacionada con la pregunta de la investigación

2. Obtener información para seguir la pista en el futuro:
    1. Dirección, número de teléfono (los números de teléfono móvil son particularmente útiles) y dirección de correo electrónico del participante
    2. Números de la seguridad social/Medicare
    3. Nombre, dirección, número de teléfono y dirección de correo electrónico de amigos o familiares cercanos que no vivan con el participante
    4. Nombre, dirección, número de teléfono y dirección de correo electrónico, del(los) médico(s)

Durante el período de seguimiento*

---

1. Contactar periódicamente con los participantes para obtener información, enviar resultados y prestar apoyo:
    1. Por teléfono: puede ser necesario llamar durante los fines de semana y por la noche
    2. Por correo: envíos repetidos por correo electrónico o con tarjetas franqueadas para su devolución
    3. Otros: boletín de noticias, regalos simbólicos
2. Para aquellos con los que no se contacta por teléfono o por correo:
    1. Contactar con amigos, familiares o médicos
    2. Solicitar direcciones de reenvío al servicio postal
    3. Buscar direcciones a través de otras fuentes públicas, como guías telefónicas e Internet
    4. Con pacientes que están en Medicare, obtener datos de altas hospitalarias de la administración de la seguridad social
    5. Determinar la situación vital con datos del Ministerio de Sanidad del estado o del National Death Registry (registro de mortalidad nacional)

En todo momento

---

1. Tratar a los participantes del estudio con amabilidad, afecto y respeto, ayudándoles a entender la pregunta de la investigación de modo que quieran participar en que el estudio tenga éxito.

---

* Esto supone que los participantes en el estudio han dado su consentimiento informado para que se recopile la información y para el contacto de seguimiento.

- En los estudios de cohortes se sigue a lo largo del tiempo a un grupo de participantes identificados al comienzo para describir la incidencia o la evolución natural de una enfermedad y descubrir los factores predictivos (factores de riesgo) de diversos resultados. La posibilidad de medir la variable predictiva antes de que se produzca la variable de respuesta establece la secuencia de los acontecimientos y controla el sesgo en esa medición.

- Los estudios de cohortes prospectivos comienzan al principio del seguimiento y pueden precisar grandes números de participantes a los que se seguirá durante períodos de tiempo prolongados. Esta última desventaja, en ocasiones, se puede superar identificando una cohorte retrospectiva en la que ya se hayan realizado las mediciones de las variables predictivas.

- El diseño de cohortes múltiples, que compara la incidencia de las variables de respuesta en cohortes que difieren en una variable predictiva («la exposición»), es útil para estudiar los efectos de exposiciones infrecuentes y de exposiciones laborales.

- Los riesgos, las oportunidades y las tasas de incidencia son tres formas de estimar la frecuencia de una variable de respuesta dicotómica durante el seguimiento; de ellas, las tasas de incidencia, que tienen en consideración las personas-tiempo de los participantes que siguen vivos y sin episodios en el estudio, forman la base de los abordajes modernos del cálculo de los cocientes de riesgo multivariados utilizando el modelo de riesgos proporcionales de Cox.

- Las inferencias sobre causa y efecto se refuerzan midiendo y ajustando todas las posibles variables de confusión que se puedan concebir. Se evitará el sesgo en la evaluación de las respuestas mediante la normalización de las determinaciones y el enmascaramiento de las personas que relacionan la variable de respuesta con los valores de la variable predictiva.

- Los puntos fuertes de un diseño de cohortes pueden debilitarse por un seguimiento incompleto de los participantes. Las pérdidas pueden reducirse al mínimo, excluyendo al comienzo a los participantes que puedan no estar disponibles para el seguimiento, recopilando información inicial que facilite su localización y permaneciendo en contacto periódicamente con todos los participantes.

Volver al principio

BIBLIOGRAFÍA

1. Andersen RE, Crespo CJ, Bartlett SJ, et al. Relationship of physical activity and television watching with body weight and level of fatness among children: results from the Third National Health and Nutrition Examination Survey. JAMA 1998;279(12):938-942. Citado aquí

2. Sargent JD, Beach ML, Adachi-Mejia AM, et al. Exposure to movie smoking: its relation to smoking initiation among US adolescents. Pediatrics 2005;116(5):1183-1191. Citado aquí

3. Jaffe HW, Bregman DJ, Selik RM. Acquired immune deficiency syndrome in the United States: the first 1,000 cases. J Infect Dis 1983;148(2):339-345. Citado aquí

4. Kalantar-Zadeh K, Abbott KC, Salahudeen AK, et al. Survival advantages of obesity in dialysis patients. Am J Clin Nutr 2005; 81: 543-554. Citado aquí

5. Zito JM, Safer DJ, DosReis S, et al. Psychotropic practice patterns for youth: a 10-year perspective. Arch Pediatr Adolesc Med 2003;157(1):17-25. Citado aquí

6. Huang Z, Hankinson SE, Colditz GA, et al. Dual effect of weight and weight gain on breast cancer risk. JAMA 1997;278:1407-1411. Citado aquí

7. Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukemia and brain tumors: a retrospective cohort study. Lancet 2012;380:499-505. Citado aquí

8. Newman TB, Liljestrand P, Jeremy RJ, et al. Outcomes of newborns with total serum bilirubin levels of 25 mg/dL or more. N Engl J Med 2006;354:1889-1900. Citado aquí

9. Escobar GJ, Liljestrand P, Hudes ES, et al. Five-year neurodevelopmental outcome of neonatal dehydration. J Pediatr 2007;151(2):127-133, 133 e1. Citado aquí

10. Wagoner JK, Archer VE, Lundin FE, et al. Radiation as the cause of lung cancer among uranium miners. N Engl J Med 1965;273:181-187. Citado aquí

CAPÍTULO 8 Diseño de estudios de casos y testigos

Thomas B. Newman
Warren S. Browner
Steven R. Cummings
Stephen B. Hulley

En el capítulo 7 se han presentado los estudios de cohortes, en los que la secuencia de las determinaciones es la misma que la cronología de causa y efecto: se miden primero las variables predictivas y después se observan las variables de respuesta durante el seguimiento. Por el contrario, en un estudio de casos y testigos, el investigador trabaja hacia atrás. Comienza eligiendo una muestra de personas con la respuesta (los casos) y otra muestra de personas sin esa respuesta (los testigos); posteriormente se comparan los niveles de las variables predictivas en las dos muestras para ver cuáles de ellas se asocian a la respuesta. Por ejemplo, en un estudio de casos y testigos se podría recoger un grupo de casos de melanoma ocular y una muestra de testigos sanos, a lo que seguiría la obtención de datos de ambos grupos sobre la exposición previa a la soldadura con arco para estimar en qué medida esta exposición afecta al riesgo de melanoma ocular. El diseño de casos y testigos es relativamente económico y tiene una eficiencia elevada para estudiar enfermedades infrecuentes.

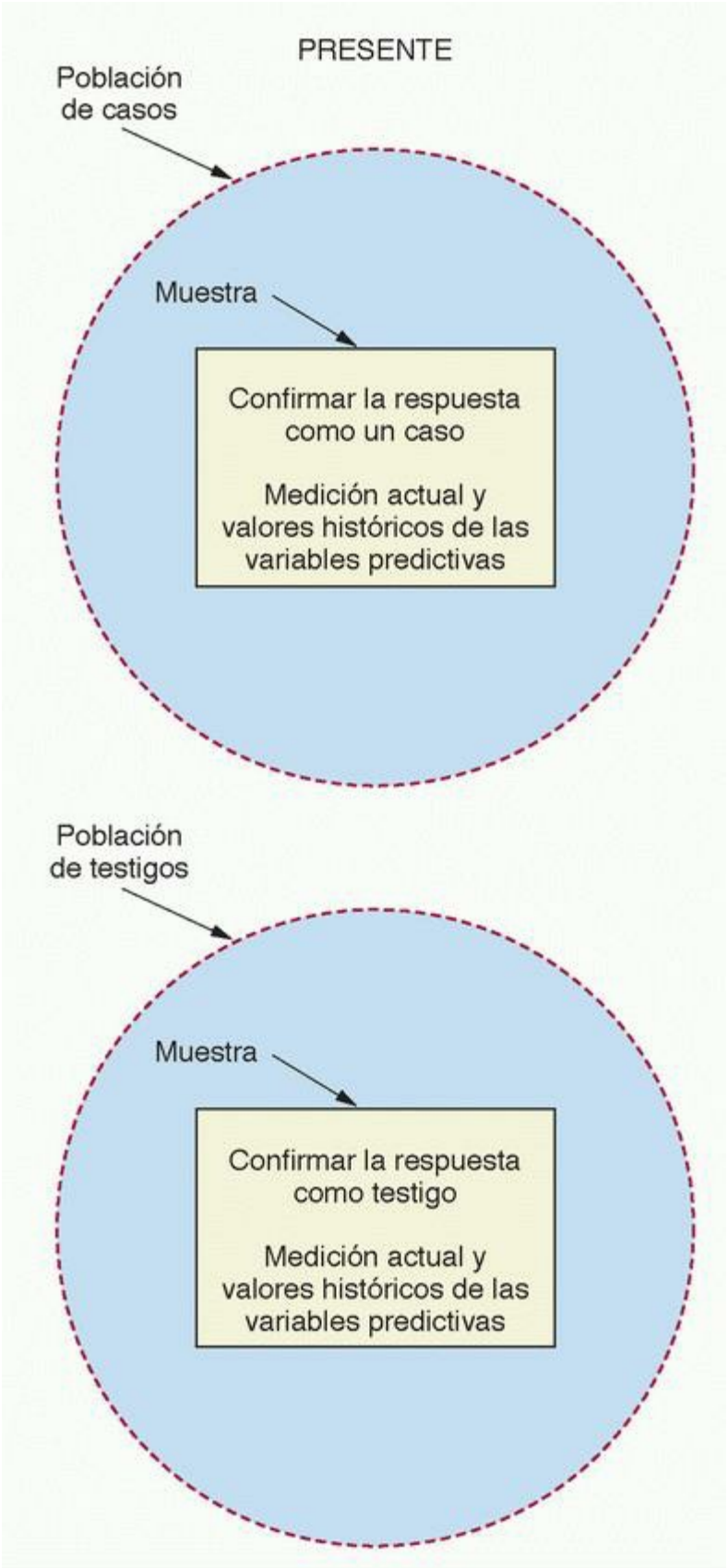
En este capítulo también se presentan diversas variaciones del diseño de casos y testigos sencillo que se ha señalado más arriba. En un diseño de casos y testigos anidado se comparan los casos incidentes anidados en un estudio de cohortes con testigos extraídos aleatoriamente del resto de la cohorte; este diseño controla el sesgo de muestreo y de medida, y ahorra dinero si las variables predictivas son mediciones costosas que se pueden realizar en muestras almacenadas o imágenes recogidas al comienzo del estudio de cohortes. Un diseño de casos y testigos de densidad de incidencia permite que los investigadores analicen las relaciones de riesgo, teniendo en consideración los cambios a lo largo del tiempo de los niveles de los factores de riesgo y las pérdidas durante el seguimiento. Y un diseño de casos y cohortes anidado permite que una muestra aleatoria de toda la cohorte actúe como testigo para diferentes conjuntos de casos. El capítulo finaliza con consejos sobre la elección de los diseños de estudios de observación que se analizan en los capítulos 7 y 8.

▪ ESTUDIOS DE CASOS Y TESTIGOS

Como la mayoría de las enfermedades son relativamente poco frecuentes, los estudios de cohortes y los estudios transversales de muestras de la población general son diseños costosos, en los que hacen falta miles de participantes para identificar factores de riesgo de una enfermedad poco habitual, como el cáncer gástrico. Como se ha señalado en el capítulo 7, una serie de casos de pacientes con la enfermedad puede identificar un factor de riesgo evidente (como el consumo de drogas por vía intravenosa en el sida), usando la información previa de la prevalencia del factor de riesgo en la población general. Sin embargo, para la mayoría de los factores de riesgo es necesario reunir un grupo de referencia, de modo que la prevalencia del factor de riesgo en las personas con la enfermedad (casos) pueda compararse con la prevalencia del factor de riesgo en las personas sin la enfermedad (testigos).

Los estudios de casos y testigos son retrospectivos (fig. 8-1). El estudio identifica un grupo de personas con la enfermedad y otro sin ella; a continuación, mira hacia atrás para encontrar diferencias en las variables predictivas que pudieran explicar por qué los casos tienen la enfermedad y los testigos no (ejemplo 8-1).

Los estudios de casos y testigos se iniciaron como estudios epidemiológicos para identificar factores de riesgo de enfermedades. Por este motivo, y como hace que sea más fácil seguir la discusión, generalmente nos referimos a los «casos» como las personas que tienen la enfermedad. Sin embargo, el diseño de casos y testigos puede usarse también para observar otras respuestas, como la incapacidad entre los que ya tienen una enfermedad. Además, cuando las respuestas no deseadas son la norma en lugar de la excepción, los casos de un estudio de casos y testigos pueden ser los pacientes, poco frecuentes, con una buena respuesta, como la recuperación de una enfermedad habitualmente mortal.



▪FIGURA 8-1. En un estudio de casos y testigos, los pasos son:
• Definir los criterios de selección e incluir una muestra de una población de casos y una segunda muestra de una población de testigos.
• Medir los valores actuales de las correspondientes variables, complementado con frecuencia con información histórica.
Herramientas de imágenes

Los estudios de casos y testigos son el «tinto de la casa» en la lista de vinos del diseño de investigación: más modestos y con algo más de riesgo que las otras selecciones, pero mucho más baratos y, a veces, sorprendentemente buenos. El diseño de un estudio de casos y testigos es difícil debido a la mayor oportunidad para que se produzca sesgo, pero hay muchos ejemplos de estudios de este tipo bien diseñados que han proporcionado resultados importantes. Entre ellos se encuentran los que establecieron los vínculos entre el consumo materno de dietilestilbestrol y el cáncer vaginal en las hijas (¡un estudio clásico que proporcionó una conclusión definitiva basándose solo en siete casos!) (1), y entre la posición de decúbito prono al dormir y el síndrome de la muerte súbita del lactante (2), un sencillo

resultado que ha salvado miles de vidas (3).

EJEMPLO 8-1. Estudio de casos y testigos

Debido a que la vitamina K intramuscular se administra sistemáticamente a los recién nacidos en Estados Unidos, un par de estudios que comunicaban una duplicación del riesgo de cáncer en la infancia entre los que habían recibido vitamina K i.m. causaron bastante impresión (4 , 5). Para investigar esta asociación, investigadores alemanes (6):

1. Seleccionaron la muestra de casos. Se seleccionaron 107 niños con leucemia del *German Childhood Cancer Registry*.

2. Seleccionaron la muestra de los testigos. Se seleccionaron 107 niños emparejados por sexo y fecha de nacimiento, y seleccionados de entre niños que vivían en la misma ciudad que el caso en el momento del diagnóstico (a partir de los registros oficiales de residencia).

3. Midieron la variable predictiva. Revisaron historias clínicas para determinar qué casos y testigos habían recibido vitamina K i.m. en el período neonatal.

Los autores encontraron que 69 de 107 casos (64 %) y 63 de 107 testigos (59 %) habían sido tratados con vitamina K i.m., para una razón de posibilidades de 1,3 (intervalo de confianza del 95 % [IC], 0,7 a 2,3). (En el apéndice 8A se expone el cálculo.) Por lo tanto, en este estudio no se confirmó la existencia de una asociación entre recibir vitamina K en el período neonatal y la posterior leucemia infantil. La estimación puntual y el límite superior del IC del 95 % dejan abierta la posibilidad de un aumento clínicamente importante de la leucemia en la población de la que se obtuvieron las muestras, pero en otros estudios y en un análisis del estudio citado usando un grupo testigo adicional no se pudo confirmar la asociación (7 , 8).

Los estudios de casos y testigos no pueden ofrecer estimaciones de la incidencia o la prevalencia de una enfermedad, porque la proporción de participantes del estudio que tiene la enfermedad se determina por la cantidad de casos y la cantidad de testigos que el investigador escoge para la muestra, en lugar de por sus proporciones en la población. Los estudios de casos y testigos proporcionan información descriptiva sobre las características de los casos y, lo que es más importante, una estimación de la solidez de la asociación entre cada una de las variables predictivas y la variable de respuesta. Estas estimaciones se presentan en forma de razón de posibilidades, que se aproxima al riesgo relativo si el riesgo de la enfermedad en los participantes expuestos y no expuestos es relativamente bajo (aproximadamente el 10 % o menos; apéndice 8B).

Puntos fuertes de los estudios de casos y testigos

Eficiencia para variables de respuesta poco habituales

Uno de los principales valores de los estudios de casos y testigos es su rápida y elevada producción de información a partir de un número relativamente bajo de participantes. Considere un estudio del efecto de la circuncisión sobre la posterior aparición de carcinoma de pene. Este cáncer es muy raro en varones circuncisos, pero también lo es en los que no lo están, cuya incidencia acumulada a lo largo de toda la vida es de alrededor del 0,16 % (9). Para realizar un estudio de cohortes con una probabilidad razonable (80 %) de detectar incluso un factor de riesgo muy importante (p. ej., un riesgo relativo de 50) se debería seguir a más de 6 000 varones durante muchos años, suponiendo que hubiera proporciones aproximadamente iguales de circuncisos y no circuncisos. Un ensayo clínico aleatorizado de la circuncisión al nacer precisaría el mismo tamaño de la muestra, pero los casos aparecerían una mediana de 67 años tras la entrada en el estudio: ¡se necesitarían tres generaciones de investigadores para realizar el seguimiento de los participantes!

Considere ahora un estudio de casos y testigos para la misma pregunta. Para la misma posibilidad de detectar el mismo riesgo relativo, solo se necesitarían 16 casos y 16 testigos (y no demasiado tiempo y esfuerzo para los investigadores). En enfermedades que son poco frecuentes o que tienen largos períodos de latencia entre la exposición y la enfermedad, los estudios de casos y testigos no solo son mucho más eficientes que otros diseños, sino que, a menudo, son la única opción posible.

Utilidad para generar hipótesis

El abordaje retrospectivo de los estudios de casos y testigos, y su capacidad de examinar un gran número de variables predictivas, hace que sean útiles para generar hipótesis sobre las causas de un nuevo brote de una enfermedad. Por ejemplo, en un estudio de casos y testigos de una epidemia de muertes por insuficiencia renal aguda en niños haitianos se observó una razón de posibilidades de 53 para la ingestión de jarabe de paracetamol de fabricación local. La investigación posterior demostró que la insuficiencia renal se debió a intoxicación con dietilenglicol, que se observó que contaminaba el jarabe de paracetamol (10), un problema que lamentablemente se ha vuelto a producir (11).

Puntos débiles de los estudios de casos y testigos

Los estudios de casos y testigos poseen grandes valores, pero también cuentan con importantes desventajas. Primero, solo se puede estudiar una variable de respuesta (la presencia o ausencia de la enfermedad, que fue el criterio para extraer las dos muestras), mientras que en los estudios de cohortes y los estudios transversales (y los estudios clínicos) se pueden estudiar diversas variables de respuesta. Segundo, como ya se ha mencionado, la información disponible en los estudios de casos y testigos es escasa: no hay ninguna forma directa de estimar la incidencia ni la prevalencia de la enfermedad, ni el riesgo atribuible ni el exceso de riesgo, salvo que el investigador también conozca la población exacta y el período temporal de los cuales se extrajeron los casos. Sin embargo, el principal punto débil de los estudios de casos y testigos es su susceptibilidad al sesgo. Este sesgo procede fundamentalmente de dos orígenes: el muestreo separado de los casos y los testigos, y la medición retrospectiva de las variables predictivas. Estos dos problemas y las estrategias para abordarlos constituyen el tema de las dos secciones siguientes.

Sesgo de muestreo y modo de controlarlo

El muestreo en un estudio de casos y testigos empieza con los casos. Idealmente, la muestra de casos debería incluir a todas las personas que hubieran presentado la enfermedad en estudio, o una selección aleatoria de dichos casos. Sin embargo, surge un problema inmediato. ¿Cómo sabemos quién ha desarrollado la enfermedad y quién no? En los estudios transversales y de cohortes, la enfermedad se busca sistemáticamente en todos los participantes del estudio, pero en los estudios de casos y testigos, la muestra de los casos debe obtenerse de pacientes en los que ya se ha diagnosticado la enfermedad y que están disponibles para el estudio. Esta muestra puede no ser representativa de todos los pacientes que desarrollan la enfermedad, porque los que no están diagnosticados, tienen un diagnóstico erróneo, no están disponibles para el estudio o fallecen tienen menos probabilidad de ser incluidos (fig. 8-2).

En general, el sesgo del muestreo es importante cuando la muestra de casos no es representativa en relación con el factor de riesgo que se está estudiando. Las enfermedades que casi siempre precisan hospitalización y que son sencillas de diagnosticar, como la fractura de cadera y las amputaciones traumáticas, pueden muestrearse con seguridad a partir de casos diagnosticados y accesibles, al menos en los países desarrollados. Por otro lado, las afecciones que pueden no llegar a ser atendidas por médicos son más difíciles de analizar en estudios de casos y testigos, debido a la selección que precede al diagnóstico. Por ejemplo, las mujeres atendidas en una consulta ginecológica con abortos espontáneos en el primer trimestre probablemente difieran de toda la población de mujeres que sufren abortos espontáneos, muchas de las cuales no solicitan asistencia médica. Por lo tanto, las mujeres con antecedentes de esterilidad estarían sobrerrepresentadas en una muestra clínica, mientras que las que tuvieran un acceso escaso a los cuidados prenatales estarían infrarrepresentadas. Si una variable predictiva de interés se asocia a la asistencia ginecológica en la población (como el uso anterior de un dispositivo intrauterino [DIU]), el muestreo de casos de la consulta podría ser una importante fuente de sesgo. Si, por otro lado, una variable predictiva no está relacionada con la asistencia ginecológica (como el grupo sanguíneo), existiría menos probabilidad de que una muestra procedente de la consulta no fuera representativa.



▪FIGURA 8-2. Algunos motivos por los que los casos de un estudio de casos y testigos pueden no ser representativos de todos los casos de la enfermedad.
Herramientas de imágenes

Aunque es importante pensar en estos aspectos, la selección de los casos a menudo está limitada a las fuentes accesibles de pacientes. La muestra de casos puede no ser totalmente representativa, pero ser lo único con lo que el investigador puede trabajar. Las decisiones difíciles a las que se enfrenta un investigador al diseñar un estudio de casos y testigos se relacionan con la tarea más abierta de seleccionar los testigos adecuados. El objetivo general es muestrear testigos de una población que habrían llegado a convertirse en casos del estudio si hubieran presentado la enfermedad. A continuación se presentan cuatro estrategias para el muestreo de los testigos:

- Testigos a partir de hospitales o consultas. Una estrategia para compensar el posible sesgo de selección, causado por obtener casos de una consulta o un hospital, es elegir testigos de los mismos centros. Por ejemplo, en un estudio del uso anterior de un DIU como factor de riesgo de aborto espontáneo, la muestra de testigos puede seleccionarse a partir de una población de mujeres que acuden por otros problemas (p. ej., vaginitis) a la misma consulta ginecológica. En comparación con una muestra aleatoria de mujeres de la misma zona, estas testigos representarían, presumiblemente, mejor a la población de mujeres que, de haber tenido un aborto espontáneo, habrían acudido a la consulta y constituirían un caso.

  Sin embargo, la selección de una muestra no representativa de testigos para compensar una muestra no representativa de casos puede ser algo problemático. Si el factor de riesgo de interés causa un problema médico para el que los testigos solicitan asistencia, la prevalencia del factor de riesgo en el grupo testigo estará falsamente elevada, lo que reduciría o invertiría la asociación entre el factor de riesgo y la respuesta. Si, por ejemplo, muchas mujeres del grupo testigo solicitaran asistencia en la consulta por una enfermedad médica asociada al uso previo de un DIU (p. ej., esterilidad por los modelos antiguos de DIU), habría un exceso de usuarias previas de DIU entre las testigos, lo que reduciría la magnitud de la asociación entre el uso previo de DIU y el aborto espontáneo en el estudio.

  Como los testigos seleccionados del hospital o la consulta muchas veces tienen enfermedades asociadas a los factores de riesgo que se estudian, los hallazgos que ofrecen estos tipos de testigos pueden llevar a error. Por lo tanto, es esencial tener en consideración si la comodidad de utilizar testigos procedentes del hospital o la consulta justifica la posible amenaza a la validez del estudio.

- Uso de una muestra de casos basada en la población. Debido a un rápido aumento del uso de los registros de enfermedades en poblaciones geográficas y en planes sanitarios, actualmente se pueden realizar estudios de casos y testigos de base poblacional

para muchas enfermedades. Los casos obtenidos de estos registros suelen ser representativos de la población general de pacientes con la enfermedad en el área de origen, lo que simplifica la elección de un grupo testigo: debe ser una muestra representativa de «no casos» procedentes de la población que abarca el registro. En el ejemplo 8-1 , el gobierno local incluyó en un registro a todos los residentes de la ciudad, lo que hizo que la selección de una muestra fuera sencilla.

Cuando se dispone de registros, los estudios de casos y testigos basados en la población son claramente los diseños más deseables. Cuando el registro de la enfermedad se acerca a la totalidad y la población que abarca se acerca a la estabilidad (no hay inmigración ni emigración), un estudio de casos y testigos basado en la población se aproxima a un estudio de casos y testigos que está anidado en un estudio de cohortes o un ensayo clínico (pág. 104), asumiendo que se puede identificar e incluir a los testigos. Estas últimas tareas son relativamente sencillas cuando la población se ha enumerado y sus historias están disponibles para los investigadores, como en el estudio de vitamina K y leucemia que se describe en el ejemplo 8-1 . Cuando no se dispone de dichas historias de registro, un abordaje que se utiliza con frecuencia es la marcación aleatoria de números de teléfono (fijos) con prefijos de la región que abarca el registro. (Cuando se seleccionan los testigos de esta forma, se deben excluir los casos que no tengan teléfono fijo.) Debido al aumento del número de hogares que solo tienen teléfono móvil, este abordaje ha llegado a ser problemático (12). Se puede realizar la marcación aleatoria, incluyendo los números de teléfono móvil, aunque se debe realizar cuidadosamente, finalizando inmediatamente si el receptor está conduciendo y evitando llamadas que pudieran suponer un coste para el receptor (13).

Sin embargo, debe reconocerse que se puede introducir sesgo siempre que se deba establecer contacto con los participantes para obtener información, porque algunos participantes (p. ej., los que no hablen inglés o los que tengan problemas de audición) pueden tener menos probabilidad de ser incluidos. Puede producirse un problema similar siempre que haga falta un consentimiento informado.

- Usar dos o más grupos testigos. Debido a que la selección de un grupo testigo puede ser demasiado difícil, particularmente cuando los casos puedan no constituir una muestra representativa de los que tienen la enfermedad, a veces es aconsejable usar dos o más grupos testigos elegidos de modos diferentes. En el estudio del Public Health Service sobre el síndrome de Reye y los fármacos (14), por ejemplo, se usaron cuatro tipos de testigos: testigos del servicio de urgencias (atendidos en el mismo servicio de urgencias que los casos), testigos ingresados (ingresados en el mismo hospital que los casos), testigos escolares (que acudían a la misma escuela o centro de día que los casos) y testigos de la comunidad (identificados por llamadas de teléfono al azar). La razón de posibilidades del uso de salicilatos en los casos en comparación con cada uno de estos grupos testigos fue, en todos los casos, de al menos 30, y fue estadísticamente muy significativa. El hallazgo constante de una intensa asociación usando grupos testigos que tendrían diferentes sesgos de muestreo refuerza la inferencia de que hay una asociación real en la población.

  Lamentablemente, pocas asociaciones tienen valores de la razón de posibilidades en modo alguno tan elevados, y los sesgos asociados a diferentes estrategias para seleccionar los testigos pueden hacer que los resultados utilizando diferentes grupos de testigos entren en conflicto mutuamente, lo que revelaría la inherente fragilidad del diseño de casos y testigos para la pregunta de la investigación que se maneja. Cuando esto sucede, el investigador debe buscar información adicional (p. ej., el motivo de consulta de los testigos procedentes de la consulta) para intentar determinar la magnitud de los posibles sesgos de cada uno de los grupos testigos (cap. 9). En cualquier caso, es mejor tener resultados incongruentes y concluir que se desconoce la respuesta, a tener tan solo un grupo testigo y extraer la conclusión errónea.

- Emparejamiento. Es un método sencillo para asegurar que los casos y los testigos sean comparables con respecto a factores importantes que están relacionados con la enfermedad, pero carecen de interés para el investigador. Hay tantos factores de riesgo y enfermedades relacionadas con la edad y el sexo, por ejemplo, que los resultados del estudio pueden no ser convincentes, salvo que los casos y los testigos sean comparables respecto a esas dos variables. Un método para evitar este problema es elegir testigos que se emparejen con los casos en estas variables predictivas constitucionales. Sin embargo, el emparejamiento tiene desventajas importantes, en concreto cuando se emparejan variables predictivas modificables, como los ingresos o la concentración de colesterol sérico. Los motivos de esto y las alternativas que generalmente se prefieren al emparejamiento se analizan en el capítulo 9.

Sesgo de medición diferencial y cómo controlarlo

El segundo punto débil importante de los estudios de casos y testigos es el riesgo de sesgo debido a un error en la medición. Esto está generado por el método retrospectivo de medición de las variables predictivas: se puede pedir a los casos y los testigos que recuerden exposiciones que se produjeron varios años antes. Lamentablemente, la memoria de las personas en relación con exposiciones previas es imperfecta. Si son imperfectas de manera similar en los casos y en los testigos, el problema se denomina error de clasificación no diferencial de la exposición y dificulta la detección de asociaciones. (En términos epidemiológicos, la razón de posibilidades está sesgada hacia el 1.) Sin embargo, es más preocupante que ser diagnosticado de una enfermedad pueda llevar a los casos a recordar o referir sus exposiciones de una manera diferente a los testigos; este error de clasificación diferencial de la exposición, denominado sesgo de

recuerdo, tiene efectos impredecibles sobre las asociaciones medidas en un estudio.

Por ejemplo, la publicidad generalizada sobre la relación entre la exposición al sol y el melanoma maligno podría llevar a los casos diagnosticados de ese cáncer a recordar su antecedente de exposición al sol de una manera diferente a los testigos. Cockburn y cols. (15) encontraron datos de este fenómeno en un inteligente estudio de gemelos discordantes en relación con el melanoma: la razón de posibilidades emparejada para tomar baños de sol en la infancia era de 2,2 (IC del 95 %: 1,0 a 4,7) cuando se preguntaba al gemelo con melanoma qué gemelo había tomado más baños de sol en la infancia, aunque era de tan solo 0,8 (0,4 a 1,8) cuando se hacía la misma pregunta al gemelo que no tenía melanoma. Sin embargo, para otras preguntas, como qué gemelo se bronceaba o se quemaba con más facilidad, no hubo datos de sesgo de recuerdo.

No puede producirse sesgo de recuerdo en un estudio de cohortes, porque se pregunta a los pacientes por las exposiciones antes del diagnóstico de la enfermedad. En un estudio de casos y testigos de melanoma maligno anidado dentro de una cohorte en la que se habían recogido varios años antes los datos de exposición al sol, se realizó un estudio directo del sesgo de recuerdo: los investigadores compararon la exposición al sol referida por los propios participantes en los casos y en los testigos tanto antes como después de que el caso fuera diagnosticado de melanoma (16). Los investigadores encontraron ciertas inexactitudes en los recuerdos de la exposición tanto en los casos como en los testigos, aunque con pocos datos de sesgo de recuerdo (16). Por lo tanto, aunque es importante tener en consideración la posibilidad de sesgo de recuerdo, no es inevitable (17).

Además de las estrategias establecidas en el capítulo 4 para controlar el sesgo en las mediciones (normalizar las definiciones operativas de las variables, escoger métodos objetivos, complementar variables clave con datos de varias fuentes, etc.), hay dos estrategias específicas para evitar el sesgo en la medición de las exposiciones en los estudios de casos y testigos:

- Usar datos registrados antes de que se produzca la respuesta. Puede que sea posible, por ejemplo, examinar las historias clínicas perinatales en un estudio de casos y testigos de uso de vitamina K intramuscular como factor de riesgo de cáncer. Esta excelente estrategia está limitada en la medida en que la información registrada sobre el factor de riesgo de interés esté disponible y sea fiable. Por ejemplo, la información sobre la administración de vitamina K no estaba a menudo en las historias clínicas, y el modo en que se trató esa información perdida afectaba a los resultados de algunos estudios de vitamina K y posterior riesgo de cáncer (8).

- Usar enmascaramiento. El método general de enmascaramiento se comentó en el capítulo 4, pero hay algunos puntos que son específicos del diseño de entrevistas en los estudios de casos y testigos. En teoría, tanto los observadores como los participantes en el estudio podrían desconocer el estado de casos y testigos de cada uno de los pacientes y el factor de riesgo que se estudia; por lo tanto, son posibles cuatro tipos de enmascaramiento (tabla 8-1).

PERSONA A LA QUE SE APLICA EL ENMASCARAMIENTO	ESTADO DEL ENMASCARAMIENTO DE CASOS Y TESTIGOS	MEDICIÓN CON ENMASCARAMIENTO DEL FACTOR DE RIESGO
Participante	Posible si tanto casos como testigos tienen enfermedades que podrían posiblemente relacionarse con el factor de riesgo	Incluir factores de riesgo «simulados» y sospechar si difieren entre los casos y los testigos Puede no funcionar si ya se ha dado publicidad al factor de riesgo de la enfermedad
Observador	Posible si los casos no se distinguen externamente de los testigos, pero sutiles signos y afirmaciones que realizan voluntariamente los participantes lo hacen difícil	Posible si el entrevistador no es el investigador, pero puede ser difícil de mantener

Idealmente, ni los participantes ni los observadores deben saber qué participantes son casos y qué participantes son testigos. En la práctica, esto muchas veces es difícil. Los participantes saben si están enfermos o no, por lo que solo se les puede ocultar su estado de caso o testigo si los testigos también presentan enfermedades que creen que podrían estar relacionadas con los factores de riesgo que se están estudiando. Los esfuerzos para que los entrevistadores desconozcan la información se ven dificultados por la naturaleza obvia de algunas enfermedades (un entrevistador puede fácilmente notar si el participante tiene ictericia o ha sufrido una laringectomía) y por los indicios que pueden obtener de las respuestas de los participantes.

El enmascaramiento de los factores de riesgo específicos que se están estudiando suele ser más fácil que el enmascaramiento del estado de casos o testigos. Un estudio de casos y testigos es, con frecuencia, el primer paso en la investigación de una enfermedad, por lo que puede que no haya tan solo un factor de riesgo de particular interés. Por ello, se pueden ocultar a los participantes del estudio y los entrevistadores las hipótesis del estudio, incluyendo preguntas «simuladas» sobre posibles factores de riesgo no asociados a la enfermedad. Por ejemplo, en un estudio de consumo de miel como factor de riesgo de botulismo del lactante se pueden incluir en la entrevista preguntas con el mismo nivel de detalle sobre el yogur y los plátanos. Este tipo de enmascaramiento no evita el sesgo diferencial, pero permite calcular si constituye un problema: si los casos comunican más exposición a la miel, pero ningún aumento de los demás alimentos, el sesgo diferencial en la medición será menos probable. Esta estrategia no funcionaría si la asociación entre tomar miel y el botulismo del lactante hubiera sido ampliamente divulgada anteriormente, o si alguno de los factores de riesgo simulados llegara a ser real.

El hecho de que el observador desconozca el estado de los participantes del estudio como casos o testigos es una estrategia particularmente buena para determinaciones de laboratorio, como los análisis de sangre y las radiografías. Es fácil el enmascaramiento en estas circunstancias, y debe realizarse siempre, simplemente pidiendo que una persona distinta a aquella que va a realizar la medición aplique una etiqueta con una identificación codificada a cada una de las muestras (o pacientes). La importancia del enmascaramiento se ilustró en 15 estudios de casos y testigos en los que se compararon determinaciones de la masa ósea de pacientes con fractura de cadera y testigos; se observaron diferencias mayores en los estudios que usaban determinaciones sin enmascaramiento que en los estudios con él (18).

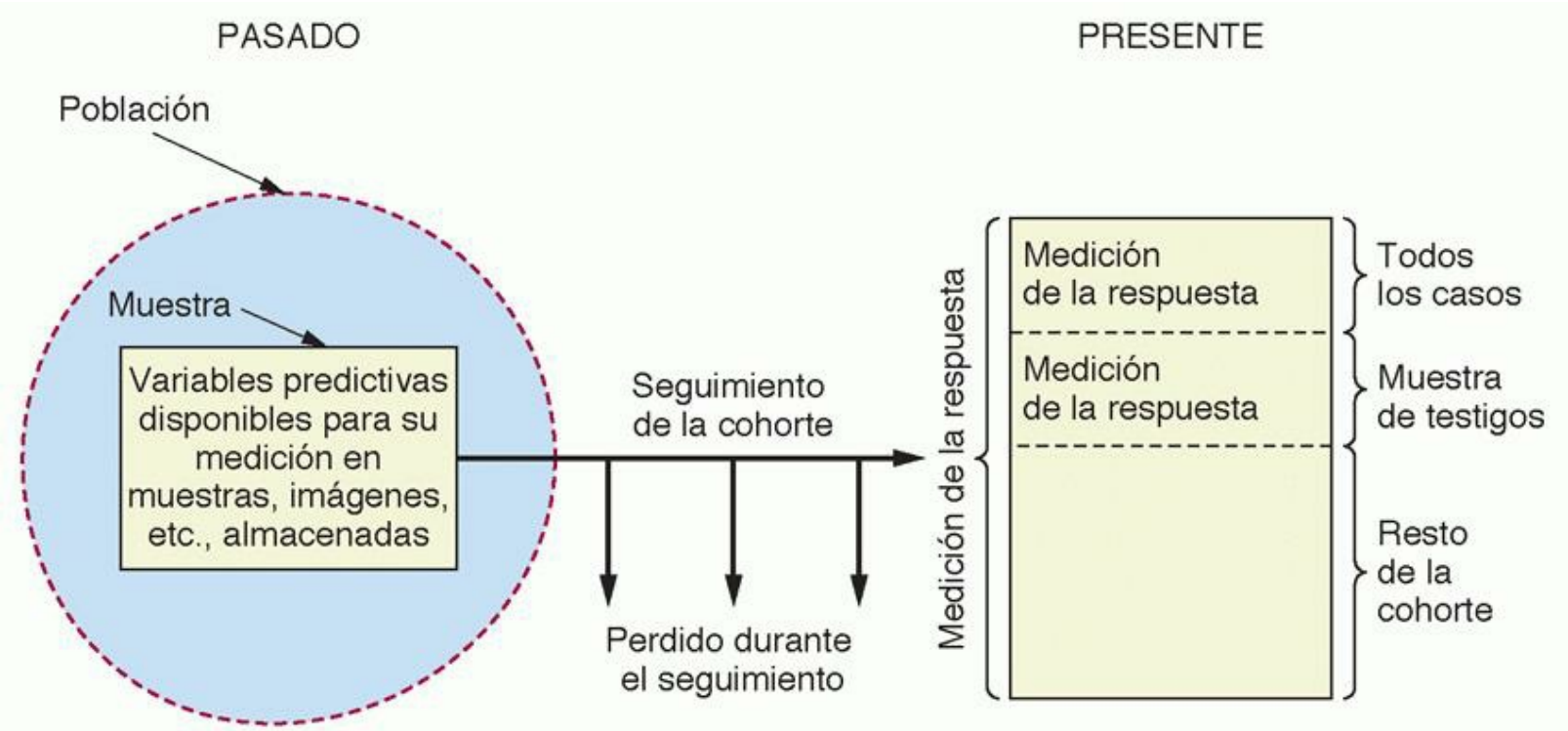
Volver al principio

▪ ESTUDIOS DE CASOS Y TESTIGOS ANIDADOS, ESTUDIOS DE CASOS Y TESTIGOS ANIDADOS CON DENSIDAD DE INCIDENCIA Y ESTUDIOS DE CASOS Y COHORTES

En un diseño de casos y testigos anidado hay un estudio de casos y testigos «anidado» dentro de una cohorte definida (fig. 8-3). El investigador puede haber definido ya la cohorte como parte de un estudio de cohortes formal, muchas veces incluyendo la conservación de muestras, imágenes y otro tipo de datos, para su análisis en el futuro después de que se produzcan las respuestas. De manera alternativa, el investigador puede diseñar un estudio de casos y testigos anidado nuevo, en una cohorte que no se haya definido previamente, y en la cual el primer paso será la definición de la cohorte.

El investigador comienza identificando una cohorte de participantes, con riesgo de que la respuesta sea suficientemente grande para que haya suficientes números de casos a fin de responder a la pregunta de la investigación, y eso ofrece la posibilidad de medir la variable de exposición, ya sea porque se han almacenado muestras o porque se dispone de historias clínicas (o pacientes) con información sobre la exposición. Como se describió en el capítulo 7, en la definición de la cohorte se debe atender a los criterios de inclusión y exclusión específicos que definan a una población en riesgo. Además, debe estar clara para cada uno de los participantes la fecha de inclusión en la cohorte. Podría tratarse de una fecha fija (p. ej., todas las personas que cumplan todos los criterios de inclusión y que estuvieran aseguradas en un plan sanitario el 1 de enero de 2008), o podría ser una fecha variable en la que comienza un período de riesgo (p. ej., la fecha de inclusión en un estudio de cohortes o la fecha del primer infarto de miocardio en un estudio de factores de riesgo de infarto de miocardio recurrente).

A continuación, el investigador describe los criterios que definen la aparición de la respuesta de interés, lo que en todos los casos se producirá después de la fecha de inclusión en la cohorte y antes del final del período de seguimiento definido. Si la respuesta es poco frecuente, el seguimiento está próximo a su fin y es suficiente una medición única de la exposición en situación inicial, entonces es sencillo. El investigador identifica a todos los participantes de la cohorte que hayan presentado la respuesta al final del seguimiento (los casos), y después selecciona una muestra aleatoria de participantes que también formaban parte de la cohorte pero que no hayan presentado la respuesta (los testigos). Después, el investigador mide las variables predictivas en los casos y en los testigos, y compara los niveles del factor de riesgo en los casos con los niveles en la muestra de testigos. Esto es un estudio de casos y testigos anidado simple (ejemplo 8-2).



▪FIGURA 8-3. Un estudio de casos y testigos anidado puede ser prospectivo o retrospectivo. Para la versión retrospectiva, los pasos son:
• Identificar una cohorte de la población con muestras, imágenes y otros datos almacenados previamente.
• Medir la variable de respuesta que distingue los casos de los testigos.
• Medir las variables predictivas en muestras, imágenes y otros datos almacenados desde que se formó la cohorte, además de otras variables, en todos los casos y en una muestra de los no casos (testigos).
Herramientas de imágenes

EJEMPLO 8-2. Diseño de casos y testigos anidado simple

Para determinar si las concentraciones elevadas de hormonas sexuales aumentaban el riesgo de cáncer de mama, Cauley (19) y cols. realizaron un estudio de casos y testigos anidado. Los pasos básicos en la realización de este estudio fueron:

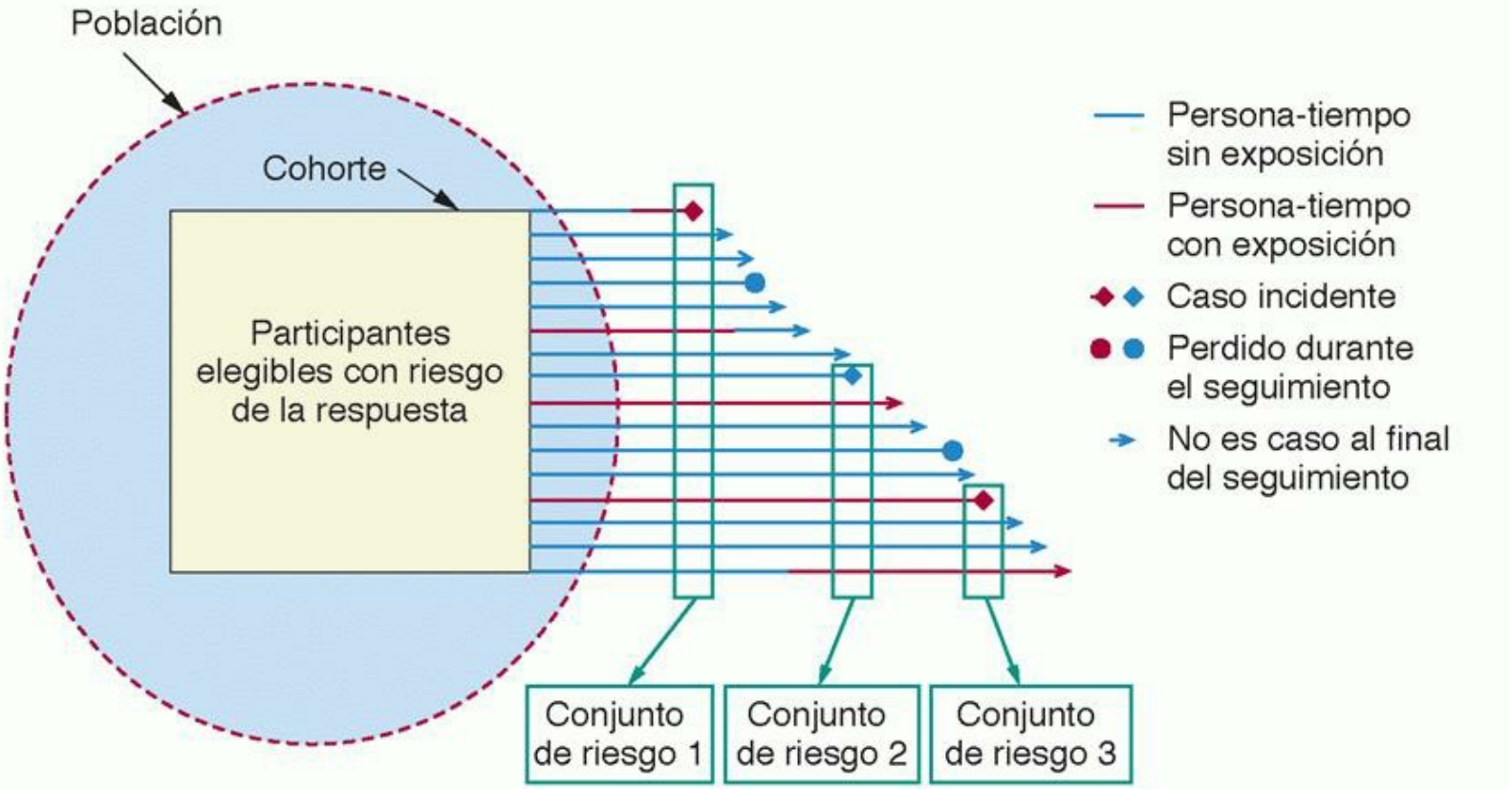
1. Identificar una cohorte. Los investigadores utilizaron la cohorte del *Study of Osteoporotic Fractures* (SOF). Fue una buena elección, porque los mismos investigadores habían extraído las muestras de suero de las participantes en esta cohorte durante la evaluación inicial y las habían almacenado congeladas a -190 °C con la esperanza de que se diseñara un estudio como este.

2. Identificar los casos al final del seguimiento. De acuerdo con las respuestas a los cuestionarios de seguimiento y una revisión de los certificados de defunción, los investigadores identificaron a 97 pacientes que habían presentado un primer episodio de cáncer de mama durante 3,2 años de seguimiento.

3. Seleccionar los testigos. Los investigadores seleccionaron una muestra aleatoria de 244 mujeres de la cohorte que no presentaron cáncer de mama durante ese período de seguimiento.

4. Medir las variables predictivas. Se midieron las concentraciones de hormonas sexuales, entre ellas el estradiol y la testosterona, en las muestras de suero congelado de la evaluación inicial de los casos y los testigos. El laboratorio desconocía si las muestras procedían de los casos o de los testigos.

Las mujeres que tenían concentraciones elevadas de estradiol o de testosterona tuvieron un aumento de tres veces del riesgo de un diagnóstico posterior de cáncer de mama en comparación con las mujeres que tenían concentraciones muy bajas de estas hormonas.



▪FIGURA 8-4. Un estudio de casos y testigos anidado con densidad de incidencia puede ser prospectivo o retrospectivo. Para la versión prospectiva, los pasos son:

• Definir los criterios de selección e incluir una cohorte de la población.
• Definir la fecha de inclusión de cada uno de los miembros de la cohorte para alinear los tiempos de seguimiento.
• Almacenar muestras, imágenes, etc., para su análisis posterior.
• Seguir la cohorte para identificar los casos y la fecha en que fueron diagnosticados.
• Obtener una muestra de uno o más testigos para cada caso a partir de los «conjuntos de riesgo», que se definen como miembros de la cohorte a los que se ha seguido durante el mismo tiempo que al caso y que no se han transformado en casos, no han muerto ni se han perdido durante el seguimiento en el momento en el que se diagnosticó al caso.
• Medir las variables predictivas en las muestras, imágenes, etc., almacenadas desde el comienzo, además de otras variables actuales, en los casos y los testigos emparejados.

Herramientas de imágenes

Si el seguimiento es variable o incompleto, o si la exposición de interés varía a lo largo del tiempo, una única medición de la exposición en la inclusión en la cohorte en los casos y una muestra aleatoria de testigos no será suficiente. En este caso es mejor diseñar un estudio de casos y testigos anidado con densidad de incidencia y obtener una muestra de los testigos de conjuntos de riesgo, definidos para cada caso como los miembros de la cohorte a los que se ha seguido durante el mismo tiempo que al caso, pero que todavía no se han convertido en casos (fig. 8-4). Como ocurre con cualquier otra forma de emparejamiento entre testigos y casos, este emparejamiento por el tiempo de seguimiento se debe tener en consideración en el análisis.

Por ejemplo, si la inclusión en la cohorte fue una fecha fija (p. ej., el 1 de enero de 2008), la muestra de testigos para un caso diagnosticado el 1 de julio de 2009 se extraería de los participantes que todavía no hubieran presentado la respuesta el 1 de julio de 2009. Si la fecha de inclusión en la cohorte fuera variable, la muestra de testigos de un caso diagnosticado 18 meses después de la inclusión se extraería de los participantes que todavía no se hubieran transformado en casos después de 18 meses de seguimiento. Dependiendo de la hipótesis de la investigación que planteara el investigador, se podrían comparar los valores de la exposición en la inclusión o en cualquier otro momento después de la inclusión entre los casos y los testigos.

Este muestreo basado en conjuntos de riesgo introduce la complejidad de que se puede seleccionar al mismo participante como testigo para un caso que se produce en fases tempranas del seguimiento y luego se puede transformar él mismo en un caso, tal vez después de que haya cambiado la intensidad de la variable de exposición. En efecto, lo que hace este diseño (con ayuda de los análisis estadísticos adecuados) es considerar de manera secuencial porciones de persona-tiempo en riesgo, utilizando para cada porción valores de las variables predictivas para predecir la aparición de casos en esa porción de persona-tiempo, con los límites de cada porción definidos por la aparición de los casos. Esto se denomina diseño de densidad de incidencia (ejemplo 8-3).

Un diseño de casos y cohortes anidado es similar al diseño de casos y testigos anidado simple, excepto que, en lugar de seleccionar testigos que no presentaron la respuesta de interés, el investigador selecciona una muestra aleatoria de todos los miembros de la cohorte, independientemente de la respuesta. Algunos participantes que formen parte de la muestra aleatoria pueden haber presentado la respuesta (el número es muy bajo cuando la respuesta es infrecuente). Una ventaja del diseño de casos y cohortes es que una única muestra aleatoria de la cohorte puede aportar los testigos para varios estudios de casos y testigos de diferentes respuestas. Además, la muestra aleatoria de la cohorte ofrece información sobre la prevalencia general de los factores de riesgo en la cohorte.

EJEMPLO 8-3. Diseño de casos y testigos anidado con «densidad de incidencia»

Para investigar la posible asociación entre el antidiabético oral pioglitazona y el cáncer vesical, investigadores de Montreal (20) realizaron un estudio de casos y testigos anidado con la base de datos *United Kingdom General Practice Research Database*, que contiene las historias clínicas de atención primaria completas de más de 10 millones de personas incluidas en más de 600 consultas generales del Reino Unido. Los pasos fueron:

1. Identificar la cohorte y el período temporal en riesgo. Los investigadores incluyeron adultos con una primera prescripción de un antidiabético oral entre el 1 de enero de 1988 y el 31 de diciembre de 2009, a los que se había seguido en la base de datos durante al menos 1 año antes de esa prescripción y que tenían al menos 40 años de edad en el momento de esa prescripción. La fecha de esta primera prescripción de un antidiabético fue la fecha de inclusión en la cohorte. Se siguió a los participantes hasta el diagnóstico de cáncer de vejiga, la muerte por cualquier causa, el final del registro en la consulta de medicina general o el final del período de estudio el 31 de diciembre de 2009, lo que se produjera primero. Se excluyó a los pacientes con antecedentes de cáncer de vejiga.

2. Identificar los casos, incluyendo las fechas de aparición. Los investigadores identificaron los casos incidentes de cáncer de vejiga utilizando *«read codes»* (códigos de lectura) (un sistema para codificar diagnósticos validados en la base de datos de investigación de medicina general [21]). Para tener en consideración la expectativa de que no cabría esperar que el efecto de la pioglitazona sobre el riesgo de cáncer fuera inmediato, excluyeron los casos que se produjeron el primer año después de la inclusión en la cohorte. Identificaron otros 376 casos de cáncer vesical.

3. Obtener una muestra de testigos de «conjuntos de riesgo» emparejados para cada uno de los casos. Los investigadores obtuvieron una muestra de hasta 20 testigos por cada caso, emparejados por el año de nacimiento, el año de inclusión en la cohorte, el sexo y la duración del seguimiento, que no habían sido diagnosticados de cáncer vesical hasta la fecha del diagnóstico del caso. El número total de testigos emparejados era de 6 699 (número medio de testigos por caso = 17,8) [1] .

4. Definir y medir las variables predictivas. La variable predictiva de interés fue recibir una prescripción de pioglitazona o rosiglitazona, otro antidiabético de la misma clase que la pioglitazona. La prescripción se debía haber realizado al menos 1 año antes de la fecha del diagnóstico del caso en el conjunto de riesgo. Se definieron cuatro niveles de exposición: prescripción solo de pioglitazona, solo de rosiglitazona, de ambas o de ninguna.

Los autores utilizaron (adecuadamente) regresión logística condicional para analizar los datos; con este método se tiene en consideración la naturaleza emparejada de los datos y, debido al muestreo del conjunto de riesgo, permite la estimación de cocientes de tasas de incidencia ajustadas (22). Encontraron cocientes de tasas de incidencia ajustadas de 1,83 (IC del 95 % 1,10 a 3,05) para el uso exclusivo de pioglitazona, 1,14 (IC del 95 % 0,78 a 1,68) para el uso exclusivo de rosiglitazona y 0,78 (IC del 95 % 0,18 a 3,29) para el uso de ambas. (El amplio intervalo de confianza del último grupo refleja un tamaño de la muestra mucho menor [$N = 2$ casos y 56 testigos].) También encontraron datos de una relación dosis-respuesta entre el uso de pioglitazona y el cáncer vesical: el cociente de tasas de incidencia ajustadas para una dosis acumulada de 28 g o más era 2,54 (1,05-6,14), p para la tendencia de dosis-respuesta = 0,03.

Puntos fuertes

Los estudios de casos y testigos anidados y de casos y cohortes son especialmente útiles para mediciones costosas en el suero y otras muestras, o para imágenes que se han archivado al comienzo del estudio y se han conservado para su análisis posterior. La realización de mediciones costosas en todos los casos y en una muestra de los testigos es mucho menos costosa que hacer las mediciones en toda la cohorte.

Este diseño conserva todas las ventajas de los estudios de cohortes, que se deben a la recogida de las variables predictivas antes de que se hayan producido las respuestas. Además, evita los posibles sesgos de los estudios de casos y testigos convencionales de que no se pueden realizar mediciones en los casos que fallecen y no se pueden extraer los casos y los testigos de poblaciones diferentes.

Puntos débiles

Estos diseños comparten ciertas desventajas con otros diseños de observación: las posibilidades de que las asociaciones observadas se

deban al efecto de variables de confusión no medidas o medidas de manera imprecisa, y de que las mediciones iniciales se puedan ver afectadas por la enfermedad preclínica silente.

Otras consideraciones

Los diseños de casos y testigos anidados y de casos y cohortes se han utilizado con menos frecuencia de lo que deberían. Un investigador que planifique estudios prospectivos extensos debe plantearse conservar muestras biológicas (p. ej., bancos de suero congelado) o almacenar imágenes o historias que sean costosas de analizar para análisis de casos y testigos anidados posteriores. Debe asegurarse de que las condiciones de almacenamiento permitan conservar durante muchos años las sustancias de interés. También puede ser útil obtener nuevas muestras o información durante el período de seguimiento, que también se podrán utilizar en las comparaciones entre casos y testigos.

Volver al principio

▪ ESTUDIOS CRUZADOS

El diseño de cruzamiento de casos es una variante del diseño de casos y testigos que es útil para estudiar los efectos a corto plazo de exposiciones intermitentes. Como los estudios de casos y testigos habituales, estos estudios retrospectivos se inician con un grupo de casos: las personas que han presentado la respuesta de interés. Sin embargo, a diferencia de los estudios de casos y testigos tradicionales, en los que se comparan las exposiciones de los casos con las exposiciones de un grupo de testigos, en los estudios cruzados cada caso actúa como su propio testigo. Las exposiciones de los casos en el momento (o justo antes) de producirse la respuesta se comparan con las exposiciones de esos mismos casos en uno o más momentos en el tiempo.

Por ejemplo, McEvoy y cols. (23) estudiaron casos que resultaron heridos en accidentes de tráfico y comunicaron tener o usar un teléfono móvil. Con los registros de la compañía telefónica, compararon el uso del móvil en los 10 min anteriores al accidente con el uso cuando los participantes iban conduciendo en el mismo momento del día anterior, 72 h antes y 7 días antes del accidente. Observaron que el uso del móvil era más probable en los 10 min previos a un accidente que en los períodos de tiempo de comparación, con una razón de posibilidades de alrededor de 4. El análisis de un estudio cruzado es como el de un estudio de casos y testigos emparejado, excepto que las exposiciones de los testigos son exposiciones del caso en diferentes períodos de tiempo en lugar de exposiciones del testigo emparejado. Esto se ilustra en el apéndice 8A, situación número 4. Los diseños de cruzamiento de casos se han utilizado en poblaciones extensas para estudiar exposiciones que varían a lo largo del tiempo, como los niveles de contaminación ambiental; se han encontrado asociaciones con el infarto de miocardio (24 , 25), las visitas al servicio de urgencias por enfermedades respiratorias (26) e incluso la mortalidad en menores de 1 año (27).

Volver al principio

▪ ELECCIÓN ENTRE DISEÑOS DE OBSERVACIÓN

En la tabla 8-2 se resumen los pros y los contras de los principales diseños de observación presentados en los dos últimos capítulos. Ya se han descrito esos temas con detalle y solo resta añadir aquí una puntualización final. Entre todos estos diseños, ninguno es el mejor y ninguno es el peor; cada uno ocupa su lugar y tiene su finalidad, dependiendo de la pregunta de investigación y de las circunstancias.

DISEÑO	VENTAJAS	INCONVENIENTES*
	Transversal	
	Duración relativamente corta	No establece la secuencia de los acontecimientos
	Un buen primer paso para un estudio de cohortes o un estudio clínico	No es posible para variables predictivas infrecuentes o variables de respuesta infrecuentes
	Ofrece la prevalencia de múltiples variables predictivas y de respuesta	No permite calcular la incidencia

	Diseños de cohortes	
Todos	Establece la secuencia de los acontecimientos Múltiples variables predictivas y de respuesta El número de sucesos de respuesta aumenta con el tiempo Permite calcular incidencia, riesgo relativo, exceso de riesgo	A menudo precisa grandes tamaños de muestra Menos factible para variables de respuesta poco frecuentes
Estudio de cohortes prospectivo	Más control sobre la selección de los participantes y las mediciones Evita el sesgo al medir variables predictivas	El período de seguimiento puede ser prolongado A menudo caro
Estudio de cohortes retrospectivo	El seguimiento es en el pasado Relativamente barato	Menos control sobre la selección de participantes y mediciones
Cohortes múltiples	Útil cuando distintas cohortes tienen exposiciones diferentes o poco frecuentes	Sesgo y confusión por muestreo en varias poblaciones
	Casos y testigos	
	Útil para respuestas poco habituales Corta duración, muestra pequeña Relativamente barato	Sesgo y confusión por muestreo en dos poblaciones Sesgo de medición diferencial Limitado a una variable de respuesta La secuencia de los acontecimientos puede no estar clara No permite calcular prevalencia, incidencia ni exceso de riesgo, salvo que esté anidado dentro de una cohorte
	Diseños híbridos	
Casos y testigos anidado	Ventajas de un diseño de cohortes retrospectivo, menos costoso si la medición de las variables predictivas es cara	Las mediciones de los factores de riesgo están sometidas a sesgos si no se habían medido previamente o si se basan en muestras o imágenes almacenadas previamente; habitualmente se necesita una cohorte definida preexistente
Casos y testigos anidado con densidad de incidencia	Permite que los investigadores analicen las relaciones de riesgo teniendo en consideración los cambios a lo largo del tiempo de los niveles de los factores de riesgo y las pérdidas durante el seguimiento	Precisa la medición de los niveles de los factores de riesgo y la incidencia de los casos a lo largo del tiempo durante el seguimiento; habitualmente precisa una cohorte definida preexistente

Casos y cohortes anidado	Igual que los estudios de casos y testigos anidados, y puede utilizar un grupo testigo único para múltiples estudios de casos y testigos con diferentes variables de respuesta	Igual que los estudios de casos y testigos anidados
Cruzamiento	Los casos sirven como sus propios testigos, reduciendo el error aleatorio y la confusión	Requiere que la exposición tenga únicamente efectos inmediatos, a corto plazo

* Todos estos diseños de observación tienen el inconveniente (en comparación con los estudios aleatorizados) de poder sufrir la influencia de variables de confusión (cap. 9).

Volver al principio

▪ RESUMEN

- En un estudio de casos y testigos se compara la prevalencia de un factor de riesgo en una muestra de participantes que tiene una variable de respuesta de interés (los casos) con la prevalencia en una muestra que no tiene dicha variable de respuesta (los testigos). Este diseño, en el que se realiza un muestreo por separado de personas con y sin la enfermedad, es relativamente barato y tiene una eficiencia elevada para estudiar enfermedades poco frecuentes.
- Un problema que surge con los estudios de casos y testigos es su susceptibilidad al sesgo de muestreo. Hay cuatro métodos para reducir este sesgo: *a)* muestrear testigos y casos del mismo (ciertamente no representativo) modo; *b)* realizar un estudio basado en la población; *c)* usar varios grupos de testigos muestreados de diferentes modos, y *d)*emparejar los casos y los testigos.
- El otro problema importante con los estudios de casos y testigos es su diseño retrospectivo, que los hace susceptibles al sesgo de medición, que afecta a casos y testigos de modo diferencial. Este sesgo puede disminuirse utilizando mediciones de la variable predictiva realizadas antes de que se produzca la variable de respuesta, y mediante el enmascaramiento de participantes y observadores.
- La mejor forma de evitar el sesgo de muestreo y de medición es diseñar un estudio de casos y testigos anidado en el que se extraen muestras aleatorias de casos y testigos de un estudio de cohortes de mayor tamaño tras su finalización. Además de controlar estos dos sesgos, las mediciones iniciales costosas en suero, imágenes, etc., se pueden realizar al final del estudio en un número relativamente pequeño de participantes en el estudio.
- El diseño de casos y testigos con densidad de incidencia permite que los investigadores analicen las relaciones de riesgo teniendo en consideración los *cambios a lo largo del tiempo* de los niveles de los *factores de riesgo* y de la disponibilidad del seguimiento.
- En el diseño de casos y cohortes anidado se utiliza una muestra aleatoria de toda la cohorte en lugar de los «no casos»; esto puede servir como grupo testigo para estudiar más de una variable de respuesta, y ofrece información directa sobre la prevalencia general de los factores de riesgo en la cohorte.
- Los estudios cruzados son una variación en el diseño de casos y testigos emparejados en el que las observaciones en dos momentos del tiempo permiten que cada caso sea su propio testigo.

Volver al principio

[1]En el capítulo 9 se señalará que el aumento de la potencia por el muestreo de más de cuatro testigos por caso es pequeño, aunque, en este caso, el coste adicional fue bajo, porque ya se disponía de los datos electrónicos. Incluso con 20 testigos por caso el abordaje de casos y testigos anidado tiene una eficiencia computacional mucho mayor que un estudio de cohortes retrospectivo.

[2]La prevalencia relativa y el exceso de prevalencia son los análogos transversales del riesgo relativo y el exceso de riesgo.

[3]Los autores realmente realizaron un análisis emparejado multivariado porque era adecuado para el diseño emparejado, pero, en este caso, la razón de posibilidades simple sin emparejamiento era casi la misma que la que se describió en el estudio.

APÉNDICE 8ACálculo de medidas de asociación
NA

1. Estudio transversal. Reijneveld (28) realizó un estudio transversal del tabaquismo materno como factor de riesgo para el cólico del lactante. Los resultados parciales fueron:

	VARIABLE DE RESPUESTA		
VARIABLE PREDICTIVA	CÓLICO DEL LACTANTE	SIN CÓLICO DEL LACTANTE	TOTAL
Madre que fuma de 15 a 50 cigarrillos/día	15 *(a)*	167 *(b)*	182 *(a + b)*
Madre no fumadora	111 *(c)*	2 477 *(d)*	2 588 *(c + d)*
Total	126 *(a + c)*	2 644 *(b + d)*	2 770 *(a + b + c + d)*

Prevalencia del cólico con madres fumadoras = $a/(a + b)$ = 15/182 = 8,2 %

Prevalencia del cólico con madres no fumadoras = $c/(c + d)$ = 111/2 588 = 4,3 %

Prevalencia total del cólico = $(a + c)/(a + b + c + d)$ = 126/2 770 = 4,5 %

Exceso de prevalencia [2] = 8,2% - 4,3% = 3,9%

2. Estudio de casos y testigos. La pegunta de investigación del ejemplo 8-1 era si existe una asociación entre la vitamina K intramuscular y el riesgo de leucemia en la infancia. Los hallazgos fueron que 69/107 casos de leucemia y 63/107 testigos habían recibido vitamina K. Una tabla de 2 × 2 con estos hallazgos sería:

	VARIABLE DE RESPUESTA: DIAGNÓSTICO	
VARIABLE PREDICTIVA: ANTECEDENTE DE FÁRMACOS	LEUCEMIA INFANTIL	TESTIGO
Vitamina K intramuscular	69 *(a)*	63 *(b)*
Sin vitamina K intramuscular	38 *(c)*	44 *(d)*
Total	107	107

Como la enfermedad (leucemia en este caso) es poco frecuente, la razón de posibilidades proporciona una buena estimación del riesgo relativo. Por lo tanto, la leucemia era aproximadamente 1,3 veces más probable después de recibir vitamina K, aunque este dato no fue estadísticamente significativo [3] .

3. Estudio de casos y testigos emparejado.

(Para ilustrar la similitud entre el análisis de un estudio de casos y testigos emparejado y estudios cruzados usaremos el mismo ejemplo para ambos.) La pregunta de la investigación es si el uso del teléfono móvil aumenta el riesgo de accidentes de tráfico entre los propietarios de estos teléfonos. Un estudio de casos y testigos emparejado podría considerar como factor de riesgo la frecuencia de uso del teléfono móvil mientras se conduce referida por los propios participantes. Los casos serían personas que habían tenido accidentes de tráfico, y se las podría comparar con testigos que no hubieran tenido accidentes, emparejados con los casos por edad, sexo y prefijo de teléfono móvil. Entonces se preguntaría a los casos y los testigos si usan alguna vez el teléfono móvil mientras conducen. (Para simplificar, para este ejemplo dividimos en dos la exposición y consideramos a las personas como «usuarios» o «no usuarios» de teléfonos móviles al conducir.) A continuación clasificamos cada par de caso/testigo según si ambos son usuarios, ninguno lo es, o el caso era usuario, pero no así el testigo, o el testigo era usuario, pero no lo era el caso. Si teníamos 300 pares, los resultados podrían ser:

	CASOS (CON LESIONES POR COLISIÓN)		
TESTIGOS EMPAREJADOS	USUARIO	NO USUARIO	TOTAL
Usuario	110	40	150
No usuario	90	60	150
Total	200	100	300

La tabla 8A-3 muestra que había 90 parejas en las que el caso usaba alguna vez el teléfono móvil mientras conducía, pero no el testigo emparejado, y 40 pares en los que el testigo emparejado era «usuario», pero el caso no lo era. Observe que esta tabla de 2 × 2 es diferente de la tabla de 2 × 2 del estudio en la pregunta 2 de la vitamina K sin emparejamiento, en el que cada celda de la tabla es el número de personas en esa celda. En la tabla de 2 × 2 para un estudio de casos y testigos *emparejado,* el número de cada celda es el número de *pares* de participantes en esa celda; el *N* total en la tabla 8A-3 es, por tanto, de 600 (300 casos y 300 testigos). La razón de posibilidades (RP) para una tabla como esta es, sencillamente, la proporción de los dos tipos de pares discordantes; en la tabla 8A-3 , la RP = 90/40 = 2,25. Esto implica que los usuarios de teléfonos móviles tenían una posibilidad mayor del doble de estar implicados en un accidente.

4. Estudio cruzado. Considere ahora el estudio cruzado de la misma pregunta. Se muestran a continuación los datos del estudio de McEvoy y cols.

	PERÍODO DE TIEMPO DE LA COLISIÓN		
SIETE DÍAS ANTES DE LA COLISIÓN	CONDUCTOR QUE USA TELÉFONO	CONDUCTOR QUE NO LO USA	TOTAL
Conductor que usa teléfono	5	6	11
Conductor que no lo usa	27	288	315
Total	32	294	326

En el estudio cruzado, cada celda de la tabla es el número de participantes, no un número de pares, pero *cada celda representa dos períodos de tiempo* para cada uno de los participantes: el período justo antes de la colisión y un período de comparación de 7 días antes. Por lo tanto, el 5 que aparece en la celda superior izquierda significa que había 5 conductores implicados en colisiones que estaban usando el móvil justo antes de chocar y que también usaron el móvil durante el período de comparación 7 días antes, mientras que el 27 situado justo debajo del 5 indica que había 27 conductores implicados en colisiones que estaban usando el móvil justo antes de la colisión, pero *no* usaban el teléfono durante el período de comparación 7 días antes. La razón de posibilidades es la proporción de los números de períodos de tiempo discordantes; en este ejemplo es de 27/6 = 4,5, lo que significa que conducir durante los períodos de uso del teléfono móvil se asocia a una posibilidad 4,5 veces mayor de accidente que conducir durante períodos en los que no se utiliza un teléfono móvil.

APÉNDICE 8BPor qué puede usarse la razón de posibilidades como estimación del riesgo relativo en un estudio de casos y testigos
NA

Los datos de un estudio de casos y testigos representan dos muestras: los casos se obtienen de una población de personas que tienen una enfermedad y los testigos de una población de personas que no la tienen. Se mide la variable predictiva (factor de riesgo), y los resultados se pueden medir en una tabla de 2 × 2 como la siguiente:

	Casos	Testigos
Presencia de factor de riesgo	*a*	*b*
Ausencia de factor de riesgo	*c*	*d*

Si esta tabla de 2 × 2 representara datos de un estudio de cohortes, la incidencia de la enfermedad en los que presentan el factor de riesgo sería $a/(a + b)$ y el riesgo relativo sería sencillamente $[a/(a + b)]/[c/(c + d)]$. Sin embargo, no es adecuado calcular la incidencia y el riesgo relativo de este modo en un estudio de casos y testigos, porque las dos muestras no se extraen de la población en las mismas proporciones. Generalmente hay cantidades aproximadamente iguales de casos y testigos en las muestras del estudio, pero muchos menos casos que testigos en la población. Más bien, el riesgo relativo en un estudio de casos y testigos puede aproximarse mediante la razón de posibilidades, calculada como el producto cruzado de la tabla de 2 × 2, ad/bc.

Este hecho extremadamente útil es difícil de captar de forma intuitiva, pero es fácil de demostrar algebraicamente. Considere la situación para toda la población, representada por a', b', c' y d'.

	Enfermedad	Sin enfermedad
Presencia de factor de riesgo	a'	b'
Ausencia de factor de riesgo	c'	d'

Aquí es adecuado calcular el riesgo de la enfermedad en personas con el factor de riesgo como $a'/(a' + b')$, el riesgo en los que no tienen el factor de riesgo como $c'/(c' + d')$, y el riesgo relativo como $[a'/(a' + b')]/[c'/(c' + d')]$. Ya se ha comentado el hecho de que $a'/(a' + b')$ no es igual a $a/(a + b)$. Sin embargo, si la enfermedad es relativamente poco frecuente tanto en los que tienen el factor de riesgo como en los que no lo tienen (que son la mayoría), entonces a' es mucho menor que b', y c' es mucho menor que d'. Esto significa que a'/b' es una buena aproximación de $a'/(a' + b')$, y que c'/d' es una buena aproximación de $c'/(c' + d')$. Por lo tanto, se puede estimar una aproximación del riesgo relativo en la población como sigue:

El último término es la razón de posibilidades de la población (literalmente, el cociente de las posibilidades de la enfermedad en los que

tienen el factor de riesgo, *a′/b′, entre* las posibilidades de la enfermedad en los que carecen de él, *c′/d′*). Esto puede reorganizarse como el producto cruzado:

Sin embargo, *a′/c′* en la población es igual a *a/c* en la muestra si los casos son representativos de todos los casos de la población (es decir, tienen la misma prevalencia del factor de riesgo). Del mismo modo, *b′/d′* es igual a *b/d* si los testigos son representativos.

Por lo tanto, los parámetros de la población en este último término pueden sustituirse por los parámetros de la muestra, y nos quedamos con el hecho de que la razón de posibilidades observada en la muestra, *ad/bc,* es una buena aproximación del riesgo relativo en la población, [*a′/(a′ + b′)*]/[*c′/(c′ + d′)*], siempre que la enfermedad sea poco frecuente.

BIBLIOGRAFÍA

1. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. N Engl J Med 1971;284(15):878-881. Citado aquí

2. Beal SM, Finch CF. An overview of retrospective case-control studies investigating the relationship between prone sleeping position and SIDS. J Paediatr Child Health 1991;27(6):334-339. Citado aquí

3. Mitchell EA, Hutchison L, Stewart AW. The continuing decline in SIDS mortality. Arch Dis Child 2007;92(7):625-626. Citado aquí

4. Golding J, Greenwood R, Birmingham K, Mott M. Childhood cancer, intramuscular vitamin K, and pethidine given during labour. BMJ 1992;305(6849):341-346. Citado aquí

5. Golding J, Paterson M, Kinlen LJ. Factors associated with childhood cancer in a national cohort study. Br J Cancer 1990;62(2):304-308. Citado aquí

6. von Kries R, Gobel U, Hachmeister A, et al. Vitamin K and childhood cancer: a population based case-control study in Lower Saxony, Germany. BMJ 1996;313(7051):199-203. Citado aquí

7. Fear NT, Roman E, Ansell P, et al. Vitamin K and childhood cancer: a report from the United Kingdom Childhood Cancer Study. Br J Cancer 2003;89(7):1228-1231. Citado aquí

8. Roman E, Fear NT, Ansell P, et al. Vitamin K and childhood cancer: analysis of individual patient data from six case-control studies. Br J Cancer 2002;86(1):63-69. Citado aquí

9. Kochen M, McCurdy S. Circumcision and the risk of cancer of the penis. A life-table analysis. Am J Dis Child 1980;134(5):484-486. Citado aquí

10. O'Brien KL, Selanikio JD, Hecdivert C, et al. Epidemic of pediatric deaths from acute renal failure caused by diethylene glycol poisoning. Acute Renal Failure Investigation Team. JAMA 1998;279(15):1175-1180. Citado aquí

11. Fatal poisoning among young children from diethylene glycol-contaminated acetaminophen - Nigeria, 2008-2009. MMWR Morb Mortal Wkly Rep 2009;58(48):1345-1347. Citado aquí

12. Puumala SE, Spector LG, Robison LL, et al. Comparability and representativeness of control groups in a case-control study of infant leukemia: a report from the Children's Oncology Group. Am J Epidemiol 2009;170(3):379-387. Citado aquí

13. Voigt LF, Schwartz SM, Doody DR, et al. Feasibility of including cellular telephone numbers in random digit dialing for epidemiologic case-control studies. Am J Epidemiol 2011;173(1):118-126. Citado aquí

14. Hurwitz ES, Barrett MJ, Bregman D, et al. Public Health Service study of Reye's syndrome and medications. Report of the main study. JAMA 1987;257(14):1905-1911. Citado aquí

15. Cockburn M, Hamilton A, Mack T. Recall bias in self-reported melanoma risk factors. Am J Epidemiol 2001;153(10):1021-1026. Citado aquí

16. Parr CL, Hjartaker A, Laake P, et al. Recall bias in melanoma risk factors and measurement error effects: a nested case-control

study within the Norwegian Women and Cancer Study. Am J Epidemiol 2009;169(3):257-266. Citado aquí

17. Gefeller O. Invited commentary: Recall bias in melanoma—much ado about almost nothing? Am J Epidemiol 2009;169(3):267-270; discussion 71-72. Citado aquí

18. Cummings SR. Are patients with hip fractures more osteoporotic? Review of the evidence. Am J Med 1985;78(3): 487-494. Citado aquí

19. Cauley JA, Lucas FL, Kuller LH, et al. Elevated serum estradiol and testosterone concentrations are associated with a high risk for breast cancer. Study of Osteoporotic Fractures Research Group. Ann Intern Med 1999;130(4 Pt 1):270-277. Citado aquí

20. Azoulay L, Yin H, Filion KB, et al. The use of pioglitazone and the risk of bladder cancer in people with type 2 diabetes: nested case-control study. BMJ 2012;344:e3645. Citado aquí

21. Hassey A, Gerrett D, Wilson A. A survey of validity and utility of electronic patient records in a general practice. BMJ 2001;322(7299):1401-1405. Citado aquí

22. Essebag V, Platt RW, Abrahamowicz M, et al. Comparison of nested case-control and survival analysis methodologies for analysis of time-dependent exposure. BMC Med Res Methodol 2005;5(1):5. Citado aquí

23. McEvoy SP, Stevenson MR, McCartt AT, et al. Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. BMJ 2005;331(7514):428. Citado aquí

24. Bhaskaran K, Hajat S, Armstrong B, et al. The effects of hourly differences in air pollution on the risk of myocardial infarction: case crossover analysis of the MINAP database. BMJ 2011;343:d5531. Citado aquí

25. Nuvolone D, Balzi D, Chini M, et al. Short-term association between ambient air pollution and risk of hospitalization for acute myocardial infarction: results of the cardiovascular risk and air pollution in Tuscany (RISCAT) study. Am J Epidemiol 2011;174(1):63-71. Citado aquí

26. Tramuto F, Cusimano R, Cerame G, et al. Urban air pollution and emergency room admissions for respiratory symptoms: a case-crossover study in Palermo, Italy. Environ Health 2011;10:31. Citado aquí

27. Scheers H, Mwalili SM, Faes C, et al. Does air pollution trigger infant mortality in Western Europe? A case-crossover study. Environ Health Perspect 2011;119(7):1017-1022. Citado aquí

28. Reijneveld SA, Brugman E, Hirasing RA. Infantile colic: maternal smoking as potential risk factor. Arch Dis Child 2000;83(4):302-303. Citado aquí

CAPÍTULO 9 Mejora de la inferencia causal en estudios de observación

Thomas B. Newman
Warren S. Browner
Stephen B. Hulley

La mayoría de los estudios de observación están diseñados para indicar que una variable predictiva puede ser una causa de una variable de respuesta, por ejemplo, que comer brócoli puede reducir el riesgo de cáncer de colon. (Son excepciones los estudios de pruebas diagnósticas y pronósticas, que se exponen en el cap. 12 .) Las asociaciones causales entre una variable predictiva y una variable de respuesta son importantes, porque pueden ofrecer conocimientos sobre la biología subyacente de una enfermedad, identificar formas de reducir o prevenir su aparición, e incluso sugerir posibles tratamientos.

Sin embargo, no todas las asociaciones que se encuentran en un estudio de observación representan relaciones de causa y efecto. De hecho, hay otras cuatro explicaciones generales para una asociación entre una variable predictiva y una variable de respuesta en un estudio observacional (tabla 9-1). Dos de ellas, el azar y el sesgo, crean asociaciones falsas entre la variable predictiva y la variable de respuesta en la muestra en estudio que no existen en la población. Otras dos, relación efecto y causa y factores de confusión, crean asociaciones reales en la población, aunque estas asociaciones no son causales en la dirección de interés. El establecimiento de que la relación de causa y efecto es la explicación más probable de una asociación exige que se demuestre que son improbables esas otras explicaciones.

Habitualmente se cuantifica el efecto causal de una variable predictiva sobre una variable de respuesta utilizando una medida de la asociación, como el cociente de riesgo o la razón de posibilidades. Por ejemplo, supóngase que en un estudio se observa que beber café se asocia a un cociente de riesgo de infarto de miocardio (IM) de 2,0. Una posibilidad (probablemente la que sería más interesante a juicio del investigador) es que beber café aumenta al doble el riesgo de IM. Sin embargo, antes de llegar a esta conclusión se deben tener en consideración y rechazar las cuatro explicaciones alternativas.

Con el azar y el sesgo, el consumo de café se asoció a un aumento al doble del riesgo de IM en el estudio, pero esa asociación realmente no está presente en la población. Por lo tanto, el azar y el sesgo son explicaciones de las asociaciones falsas (es decir, no reales) en un estudio.

Las otras dos alternativas (efecto y causa, y confusión) son fenómenos biológicos verdaderos, lo que significa que las personas que beben café en la población realmente tienen el doble de riesgo de IM. Sin embargo, este aumento del riesgo no se debe a una relación de causa y efecto. En una situación, la asociación se debe a efecto y causa: tener un IM hace que las personas beban más café. (Esto es simplemente la relación de causa y efecto a la inversa.) Se produce la última posibilidad, la confusión, cuando un tercer factor, como el tipo de personalidad, lleva tanto al consumo de café como al IM.

En el resto de este capítulo se analizarán estrategias para estimar y minimizar la probabilidad de estas cuatro explicaciones alternativas para encontrar una asociación en un estudio de observación. Estas estrategias se pueden utilizar cuando se diseña un estudio o cuando se analizan sus resultados. Aunque este libro insiste en el diseño de la investigación, el conocimiento de las opciones analíticas puede influir en la elección del diseño, por lo que en este capítulo se abordarán los dos temas.

▪ ASOCIACIONES FALSAS DEBIDAS AL AZAR

Suponga que, en realidad, no existe ninguna asociación entre beber café y el IM en los miembros de la población, el 45 % de los cuales beben café. Si fuéramos a seleccionar 20 casos con IM y 20 testigos, podríamos esperar que aproximadamente 9 personas de cada grupo (el 45 % de 20) bebieran café.

EXPLICACIÓN	TIPO DE ASOCIACIÓN	¿QUÉ OCURRE REALMENTE EN LA POBLACIÓN?	MODELO CAUSAL
1. Azar (error aleatorio)	Falsa	Consumir café e IM no se relacionan	-
2. Sesgo (error sistemático)	Falsa	Consumir café e IM no se relacionan	-

3. Efecto y causa	Real	El IM es una causa del consumo de café	IM → Consumir café
4. Confusión	Real	Un tercer factor conduce al consumo de café y el IM	
5. Causa y efecto	Real	Consumir café es una causa de IM	Consumir café → IM

Sin embargo, *solo por azar*, podríamos incluir a 12 bebedores de café entre los 20 casos de IM, y solo 6 en los 20 testigos. Si esto ocurriera, en nuestro estudio observaríamos una asociación falsa entre el consumo de café y el IM.

El azar se denomina en ocasiones error aleatorio, porque no tiene ninguna explicación subyacente. Cuando una asociación debida a un error aleatorio es estadísticamente significativa, se conoce como error de tipo I (cap. 5).

Existen estrategias para reducir el error aleatorio tanto en la fase de diseño de la investigación como en la fase de análisis (tabla 9-2). Las e*strategias de diseño*, como aumentar la precisión de las medidas e incrementar el tamaño de la muestra, se analizan en los capítulos 4 y 6, respectivamente. La *estrategia de análisis* de calcular valores de***p*** e intervalos de confianza ayuda al investigador a cuantificar la magnitud de la asociación observada en comparación con lo que podría haber sucedido solo por azar. Por ejemplo, un valor de *p* de 0,10 indica que el azar por sí solo podría causar una diferencia al menos tan grande como la que observaron los investigadores en aproximadamente el 10 % de las ocasiones. Los intervalos de confianza, que son incluso más útiles que los valores de *p*, muestran los posibles valores de los estadísticos que describen una asociación que caen dentro del intervalo del error aleatorio estimado en el estudio.

TIPO DE ASOCIACIÓN FALSA	**FASE DE DISEÑO (CÓMO EVITAR LA EXPLICACIÓN ALTERNATIVA)**	**FASE DE ANÁLISIS (CÓMO EVALUAR LA EXPLICACIÓN ALTERNATIVA)**
Azar (debido a error aleatorio)	Aumentar el tamaño de la muestra y otras estrategias para aumentar la precisión (caps. 4 y 6)	Calcular valores de *p* e intervalos de confianza e interpretarlos en el contexto de los datos anteriores (cap. 5)
Sesgo (debido a error sistemático)	Considerar cuidadosamente las posibles consecuencias de cada diferencia entre la pregunta de la investigación y el plan del estudio (v. fig. 1-6); alterar el plan del estudio si es necesario	Comprobar la congruencia con otros estudios (especialmente los que utilizan diseños diferentes)
	Recoger datos adicionales que permitan la evaluación de la magnitud de los posibles sesgos	Analizar datos adicionales para ver si realmente se han producido posibles sesgos
	No utilizar como criterios de inclusión o variables de emparejamiento variables afectadas por la variable predictiva de interés	No controlar variables afectadas por la variable predictiva

Volver al principio

▪ ASOCIACIONES FALSAS DEBIDAS AL SESGO

Se han identificado muchos tipos de sesgo (error sistemático), y el abordaje de algunos de ellos es un tema importante de este libro. Junto con las estrategias específicas que se describen en los capítulos 3, 4, 7 y 8, ahora se añade un abordaje general para reducir la

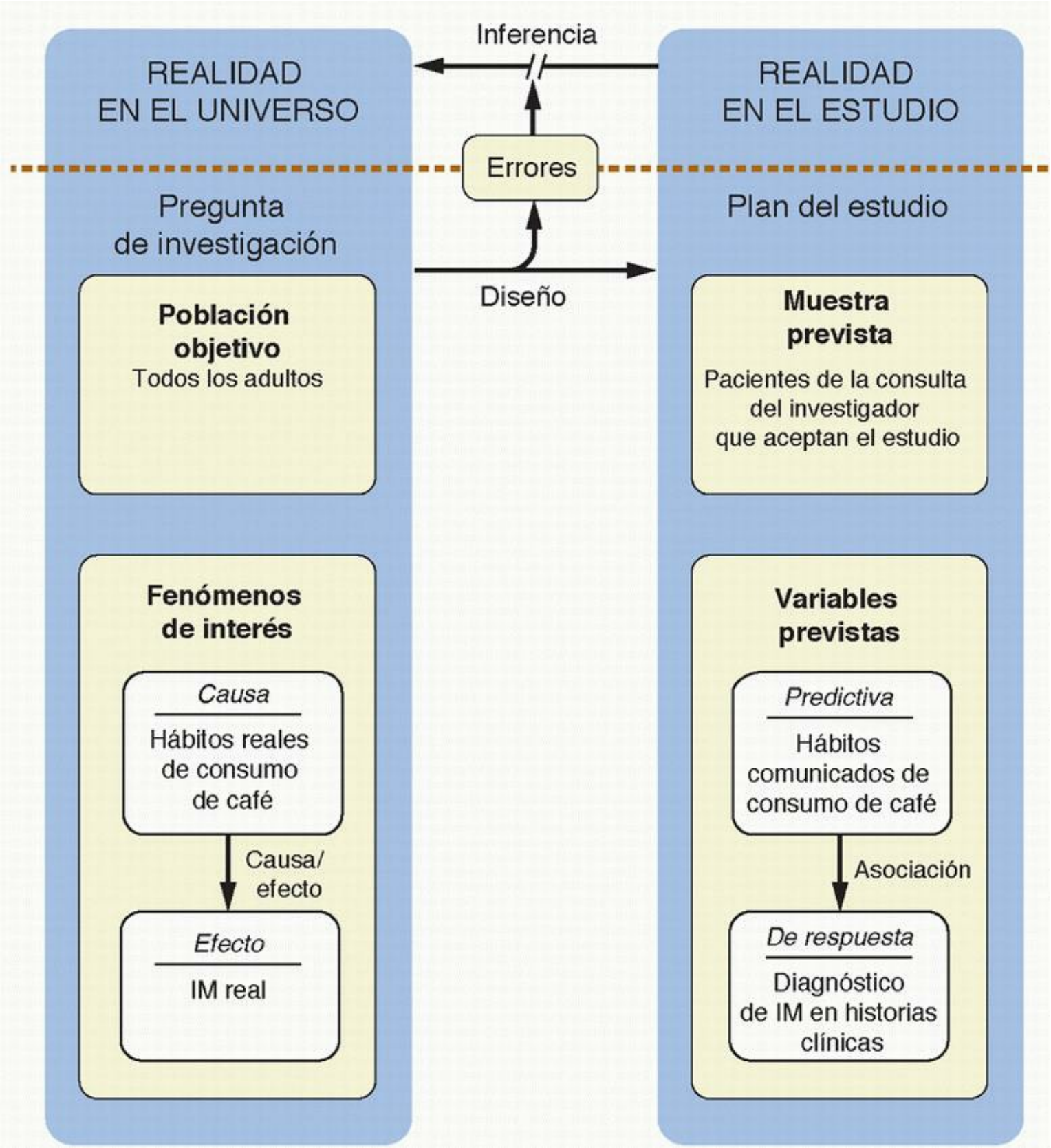
probabilidad de sesgo.

Minimización del sesgo

Como se ha señalado en el capítulo 1, casi siempre hay diferencias entre la pregunta de la investigación original y la que realmente se responde en el estudio. Estas diferencias reflejan los compromisos que se hicieron para que el estudio fuera viable, además de errores en el diseño o la ejecución del estudio. Se produce sesgo cuando esas diferencias hacen que la respuesta que ofrece el estudio difiera de la respuesta correcta a la pregunta de la investigación. Se dispone de estrategias para minimizar el sesgo en las fases de diseño y de análisis de la investigación (v. tabla 9-2):

- Fase de diseño. Comience escribiendo la pregunta de la investigación al lado del plan del estudio, como en la figura 9-1 . Después medite sobre los tres problemas siguientes en relación con la pregunta de la investigación:
    - ¿Las muestras de los participantes del estudio (p. ej., casos y testigos, o participantes expuestos y no expuestos) representan a la(s) población(es) de interés?
    - ¿Representan las mediciones de las variables predictivas los factores predictivos de interés?
    - ¿Representan las mediciones de las variables de respuesta las respuestas de interés?

  Para cada pregunta respondida «No» o «Puede que no», considere si el sesgo se aplica de modo similar a uno o a los dos grupos estudiados (p. ej., casos y testigos, o expuestos y no expuestos), y si es probable que el sesgo sea suficientemente grande como para afectar a la respuesta a la pregunta de la investigación.



▪FIGURA 9-1. Reducción al mínimo del sesgo por la consideración cuidadosa de las diferencias entre la pregunta de la investigación y el plan del estudio.
Herramientas de imágenes

Para ilustrar esto con nuestro ejemplo del café y el IM, considérese un estudio de casos y testigos en el que la muestra de los testigos se obtenga de pacientes hospitalizados por enfermedades distintas al IM. Si muchos de estos pacientes tienen enfermedades crónicas que hicieron que redujeran su consumo de café, la muestra de los testigos no representará a la población objetivo de la que surgen los casos con IM: habrá una escasez de bebedores de café. Y si el espasmo esofágico, que puede empeorar por el café, se diagnostica erróneamente como IM, podría encontrarse una falsa asociación entre el café y el IM, porque la respuesta medida (diagnóstico de IM) no representaba exactamente la variable de respuesta de interés (IM real).

El siguiente paso será pensar en posibles estrategias para evitar cada uno de los posibles tipos de sesgo, como seleccionar más de

un grupo testigo en un estudio de casos y testigos (cap. 8) o las estrategias para reducir el sesgo de medición descritas en el capítulo 4. En cualquier caso, hacen falta juicios sobre la probabilidad de los sesgos y la facilidad con la que se podrían evitar con cambios en el plan del estudio. Si el sesgo se puede evitar fácilmente, revise el plan del estudio y plantéese las tres preguntas de nuevo. Si el sesgo no se puede evitar fácilmente, decida si sigue mereciendo la pena realizar el estudio, determinando la probabilidad del posible sesgo y el grado en que distorsionará la asociación que está intentando estimar.

Los posibles sesgos pueden ser inevitables, o puede ser costoso evitarlos, o puede haber incertidumbre en la medida en la que serán problemáticos. En cualquier caso, el investigador debe plantearse diseñar el estudio para obtener datos adicionales que permitan evaluar la gravedad de los sesgos. Por ejemplo, si el investigador tiene la preocupación de que los casos de un estudio de cáncer pancreático pueden notificar en exceso exposiciones recientes a productos químicos (tal vez porque estos pacientes están buscando desesperadamente una explicación de por qué tienen cáncer pancreático), también se les podría preguntar por exposiciones (¡como el consumo de café!) que en estudios previos no se ha demostrado que tengan efecto sobre el riesgo de cáncer pancreático. Si el investigador está preocupado de que un cuestionario no detecte con exactitud el consumo de café (tal vez debido al mal planteamiento de las preguntas), podría asignar un entrevistador con enmascaramiento para que preguntara a un grupo de casos y testigos a fin de determinar la concordancia con las respuestas que han dado al cuestionario. De manera similar, si se plantea la preocupación de que, en lugar de producir IM, el café aumenta la supervivencia en pacientes con IM (lo que llevaría a que los bebedores de café estuvieran sobrerrepresentados en una muestra de supervivientes a un IM), el investigador podría identificar a los pacientes con IM que han muerto y entrevistar a sus cónyuges supervivientes sobre sus hábitos previos de consumo de café.

- Fase de análisis. Una vez que se han recogido los datos, el objetivo pasa de minimizar el sesgo a evaluar su probable gravedad. El primer paso es analizar datos que se han recogido para esa finalidad. Por ejemplo, el investigador que prevé un recuerdo imperfecto de los hábitos de consumo de café puede haber incluido preguntas sobre la seguridad que tienen los casos y los testigos en sus respuestas. Se podría examinar la asociación entre el consumo de café y el IM después de estratificar la certidumbre sobre el consumo de café, para ver si la asociación es más fuerte en los que están más seguros de su antecedente de exposición.

El investigador también puede contemplar los resultados de otros estudios. Si las conclusiones son compatibles, es menos probable que la asociación se deba al sesgo. Esto es especialmente cierto si en los demás estudios se han usado diferentes diseños y, por tanto, no es probable que compartan los mismos sesgos. En muchas situaciones, los posibles sesgos no llegan a ser un problema importante. La decisión sobre el interés con que se debe buscar información adicional y sobre cuál es la mejor forma de comentar estos aspectos cuando se elabore el informe del estudio son temas de opinión para los que resulta útil pedir consejos a compañeros.

Volver al principio

▪ ASOCIACIONES REALES APARTE DE LAS DE CAUSA Y EFECTO

Además del azar y el sesgo, deben considerarse los dos tipos de asociaciones que son reales pero no representan una relación de causa y efecto (tabla 9-3).

TIPO DE ASOCIACIÓN REAL	FASE DE DISEÑO (CÓMO EVITAR LA EXPLICACIÓN RIVAL)	FASE DE ANÁLISIS (CÓMO EVALUAR LA EXPLICACIÓN RIVAL)
Efecto y causa (la respuesta es realmente la causa de la variable predictiva)	Realizar un estudio longitudinal para descubrir qué se produjo primero Obtener datos de la secuencia histórica de las variables (Solución final: realizar un ensayo aleatorizado)	Considerar la verosimilitud biológica Comparar la fuerza de la asociación inmediatamente después de la exposición a la variable predictiva con la fuerza en momentos posteriores Considerar hallazgos de otros estudios con diferentes diseños
Confusión (otra variable produce tanto la variable predictiva como la variable	Véase tabla 9-4	Véase tabla 9-5

de respuesta)

Efecto y causa

Una posibilidad es que se invierta la secuencia: la variable de respuesta ha causado la variable predictiva. Esta posibilidad a menudo es un problema en los estudios transversales y de casos y testigos: ¿un estilo de vida sedentario produce obesidad, o es al revés? La relación de efecto y causa también puede ser un problema en estudios cruzados. Por ejemplo, en el estudio de uso de teléfonos móviles y accidentes de tráfico descrito en el capítulo 8 (1), un accidente de tráfico podría llevar al conductor a que hiciera una llamada de teléfono en la que comunicara el accidente, en lugar de que el accidente hubiera estado producido porque el conductor no prestaba atención. Para orientar esta posibilidad, los investigadores preguntaron a los conductores sobre el uso del móvil antes y después de la colisión, y comprobaron las respuestas mediante registros de llamadas telefónicas.

La relación de efecto y causa es, con menos frecuencia, un problema en los estudios de cohortes para determinar la causa de las enfermedades, porque las determinaciones de los factores de riesgo pueden realizarse en los participantes que todavía no tienen la enfermedad. Sin embargo, incluso en estos estudios puede haber relación de efecto y causa si la enfermedad tiene un largo período de latencia y no pueden identificarse al principio los pacientes que sufren la enfermedad subclínica. Por ejemplo, la diabetes de tipo II se asocia a un riesgo posterior de sufrir cáncer de páncreas. Alguna de estas asociaciones podría muy bien ser del tipo de efecto y causa, porque el cáncer de páncreas podría afectar a las células insulares pancreáticas que segregan insulina, lo que produciría diabetes. En consonancia con la relación de efecto y causa, el riesgo de cáncer pancreático es máximo inmediatamente después del diagnóstico de diabetes (2). La asociación disminuye con la duración de la diabetes, aunque persiste cierto grado de asociación incluso 4 años o más después del inicio de la diabetes (2 , 3 , 4), lo que indica que al menos parte de la relación puede ser de causa y efecto.

Este ejemplo ilustra un método general para descartar la relación de efecto y causa: buscar una disminución de la asociación al aumentar el tiempo transcurrido entre la supuesta causa y su efecto. Un segundo abordaje es evaluar la verosimilitud biológica de la relación de efecto y causa en comparación con la relación de causa y efecto. En este ejemplo, la relación de efecto y causa era creíble porque el cáncer pancreático podría dañar el páncreas, aunque la observación de que haber tenido diabetes durante más de 10 años se asocia a aumento del riesgo de otros diversos cánceres además del cáncer pancreático (4), incrementa la verosimilitud biológica de que la diabetes produce cáncer de páncreas, en lugar de ser tan solo uno de sus efectos.

Confusión

La otra explicación de la tabla 9-3 es la confusión, algo que sucede cuando un tercer factor es una causa real de la variable de respuesta, y la variable predictiva de interés se asocia a ese tercer factor, aunque no es una causa de él. Por ejemplo, si determinados rasgos de la personalidad hacen que las personas consuman más café y también que tengan mayor riesgo de IM, estos rasgos de la personalidad introducirán confusión en la asociación entre el café y el IM. Si esta es la explicación real, entonces la asociación entre el café y el IM no representa una relación de causa y efecto, aunque es perfectamente real: el consumo de café es un «espectador inocente» en relación con la causalidad.

Para que sea una variable de confusión, una variable se debe asociar con la variable predictiva de interés, y también debe ser una causa de la variable de respuesta. La confusión puede ser aún más complicada, y en ocasiones está implicado otro factor adicional. Por ejemplo, el entorno laboral podría hacer que las personas consumieran café y fumaran cigarrillos, lo cual es un factor de riesgo de IM. En el apéndice 9A se presenta un ejemplo numérico de cómo diferencias en el consumo de cigarrillos podrían llevar a una asociación aparente entre el consumo de café y el IM.

¿Qué ocurriría si el consumo de café llevara al tabaquismo y el tabaquismo produjera IM? En ese caso se consideraría que el tabaquismo es un mediador de la asociación (causal) entre el consumo de café y el IM, no un factor de confusión. En general, es mejor no controlar los factores que están a lo largo de la vía causal entre una variable predictiva y una variable de respuesta.

Aparte del sesgo, la confusión es, a menudo, la única explicación alternativa posible a la relación de causa y efecto, y la más importante que hay que intentar descartar. También es el mayor reto; gran parte del resto del capítulo se destina a las estrategias para abordar las variables de confusión. Sin embargo, se debe señalar que todas estas estrategias suponen la realización de juicios, y que no hay ninguna cantidad de sofisticación epidemiológica o estadística que pueda sustituir al conocimiento de la biología subyacente.

Volver al principio

▪ ABORDAJE DE LAS VARIABLES DE CONFUSIÓN EN LA FASE DE DISEÑO

La mayor parte de las estrategias para abordar las variables de confusión requiere que un investigador las mida, por lo que es útil comenzar enumerando las variables (como edad y sexo) que pueden estar asociadas a la variable predictiva y que también sean una causa de la variable de respuesta. El investigador deberá entonces escoger entre estrategias de diseño y análisis para controlar la influencia de estas posibles variables de confusión.

Las dos primeras estrategias de la fase de diseño (tabla 9-4), especificación y emparejamiento, suponen cambios en el esquema del muestreo. Se pueden obtener las muestras de los casos y los testigos (en un estudio de casos y testigos) o de los participantes expuestos y no expuestos (en un estudio de cohortes), de modo que presenten valores comparables de la variable de confusión. Esto elimina esta última como explicación de cualquier asociación que se observe entre la variable predictiva y la de respuesta. Una tercera estrategia de la fase de diseño, la utilización de diseños de estudio oportunistas, solo puede aplicarse para seleccionar preguntas de investigación para las que se den las condiciones correctas. Sin embargo, cuando pueden aplicarse, estos diseños se parecen a ensayos aleatorizados, en cuanto a su capacidad de reducir o eliminar la confusión, que introducen no solo variables que se han medido, sino también variables que no se han medido.

Especificación

La estrategia más sencilla consiste en diseñar criterios de inclusión que especifiquen un valor de la posible variable de confusión y excluyan a cualquiera que tenga un valor diferente. Por ejemplo, el investigador que estudia el café y el IM podría especificar que solo se incluyan en el estudio los no fumadores. Si se observara entonces una asociación entre el café y el IM, evidentemente no podría deberse al tabaquismo.

La especificación es una estrategia eficaz, pero, como todas las restricciones en el esquema de muestreo, también presenta inconvenientes. En primer lugar, incluso si el café no es una causa de IM en los no fumadores, sí puede causarlo en los que fuman. Este fenómeno, un efecto del café sobre el IM que es diferente en los fumadores y en los no fumadores, se denomina modificación del efecto (también conocido como interacción); véase el apéndice 9A. Por ello, la especificación limita la posibilidad de generalizar la información disponible de un estudio, en este caso, comprometiendo nuestra capacidad de generalizar a los fumadores. Una segunda desventaja es que, si el tabaquismo es muy frecuente en los pacientes disponibles para el estudio, puede que el investigador no sea capaz de conseguir una muestra suficientemente grande de no fumadores. Estos problemas pueden llegar a ser importantes si la especificación se utiliza para controlar demasiadas variables de confusión o para controlarlas demasiado estrictamente. El tamaño de la muestra y la posibilidad de generalización serían problemas importantes si se limitara un estudio a varones no fumadores de 70 a 74 años de edad con un nivel de ingresos bajo.

ESTRATEGIA	VENTAJAS	INCONVENIENTES
Especificación	• Fácilmente comprensible • Delimita la muestra de participantes para la pregunta de la investigación	• Limita la generalización y el tamaño de la muestra
Emparejamiento	• Puede eliminar la influencia de variables de confusión importantes, como la edad y el sexo • Puede eliminar la influencia de variables de confusión difíciles de medir • Puede aumentar la potencia equilibrando el número de casos y testigos en cada estrato • Puede ser una conveniencia de muestreo, facilitando la selección de testigos en un estudio de casos y	• Puede requerir tiempo y ser costoso; puede ser menos eficaz que aumentar el número de participantes • Debe tomarse la decisión al principio del estudio, y puede tener un efecto irreversible sobre el análisis • Requiere una decisión precoz sobre qué variables son predictivas y cuáles de confusión • Elimina la opción de estudiar variables emparejadas como predictivas o variables de intervención • Requiere un análisis emparejado • Crea el peligro de emparejamiento excesivo (emparejamiento en un factor que no es variable de confusión, con lo que reduce la potencia)

	testigos	• Solo factible en estudios de casos y testigos y de múltiples cohortes
Diseños de estudios «oportunistas»	• Puede proporcionar una gran solidez de inferencia causal • Puede ser una alternativa buena y menos costosa al ensayo clínico aleatorizado	• Solo es posible en determinadas circunstancias en las que la variable predictiva se asigna de forma aleatoria o prácticamente aleatoria, y existe una variable instrumental

Emparejamiento

En un estudio de casos y testigos, el emparejamiento se puede utilizar para evitar la confusión, seleccionando los mismos valores (emparejados) de las variables de confusión. El emparejamiento y la especificación evitan la confusión al permitir la comparación solo de casos y testigos que comparten niveles comparables de la variable de confusión. Sin embargo, el emparejamiento difiere de la especificación en que se mantiene la posibilidad de generalizar, porque pueden estudiarse participantes de todos los niveles de la variable de confusión.

El emparejamiento suele realizarse individualmente (emparejamiento por parejas). Para controlar el tabaquismo en un estudio en el que beber café era una variable predictiva del IM, por ejemplo, se podría emparejar cada caso (un paciente con un IM) individualmente con uno o más testigos que fumaran aproximadamente lo mismo que el caso (p. ej., 10 a 20 cigarrillos/día). El consumo de café de cada caso podría compararse con el consumo de café del(los) control(es) emparejado(s).

Un enfoque alternativo es emparejar por grupos (emparejamiento por frecuencias). Para cada grado de tabaquismo podría contarse el número de casos con ese consumo, y se selecciona un número adecuado de testigos con el mismo nivel de tabaquismo. Si el estudio requería dos testigos por caso y había 20 casos que fumaban 10 a 20 cigarrillos/día, el investigador seleccionaría 40 testigos que fumaran esa cantidad, emparejándolos como grupo con los 20 casos.

El emparejamiento se usa fundamentalmente en los estudios de casos y testigos, pero también puede utilizarse en diseños de cohortes múltiples. Por ejemplo, para investigar los efectos de la participación en la Guerra del Golfo de 1990 a 1991 sobre la fecundidad en veteranos varones, Maconochie y cols. (5) compararon varones desplegados en la región del Golfo durante la guerra con varones no desplegados, pero que estaban emparejados por el número de años en el ejército, la edad y el entrenamiento para el despliegue, entre otros factores. Se observó que había un riesgo ligeramente mayor de esterilidad comunicada (RP ~ 1,5) y un mayor tiempo hasta la concepción entre los veteranos de la Guerra del Golfo.

Ventajas del emparejamiento (v. tabla 9-4)

- El emparejamiento es una forma eficaz de evitar la confusión por factores constitucionales, como la edad, el sexo y la raza, que son potentes determinantes de la respuesta, no susceptibles de intervención y con poca probabilidad de ser un intermediario en una vía causal.

- El emparejamiento puede usarse para controlar variables de confusión que no pueden medirse ni controlarse de ningún otro modo. Por ejemplo, emparejar hermanos (o, mejor aún, gemelos) entre sí puede controlar un amplio espectro de factores genéticos y familiares que serían imposibles de medir. El emparejamiento por centro médico en un estudio multicéntrico puede controlar diferencias inespecíficas entre las poblaciones o el personal en centros con dispersión geográfica.

- El emparejamiento puede aumentar la precisión de las comparaciones entre grupos (y, por tanto, la potencia del estudio para encontrar una asociación real), equilibrando el *número* de casos y testigos en todos los niveles de la variable de confusión. Esto puede ser importante si el número disponible de casos está limitado o si el coste de estudiar a los participantes es elevado. No obstante, el efecto del emparejamiento sobre la precisión es escaso y no siempre favorable (v. «emparejamiento excesivo», pág. 125). En general, el deseo de aumentar la precisión es un motivo menos importante para el emparejamiento que la necesidad de controlar la confusión.

- Finalmente, el emparejamiento puede usarse fundamentalmente como un método conveniente de muestreo, para reducir un número de posibles testigos imposiblemente grande. Por ejemplo, en un estudio de consumo de marihuana como factor de riesgo de tumores de células germinales testiculares, los investigadores pidieron a los casos (hombres con tumores testiculares) que propusieran a amigos de una edad similar sin tumores que participaran en el grupo testigo (6). Esta conveniencia, no obstante,

también conlleva el riesgo de emparejamiento excesivo.

Desventajas del emparejamiento (v. tabla 9-4)

- El emparejamiento requiere tiempo y gastos adicionales para identificar una pareja para cada participante. En estudios de casos y testigos, cuantos más criterios de emparejamiento haya, mayor será el número de testigos que pueden buscarse para emparejarlos con cada caso. El posible aumento de la potencia estadística por el emparejamiento deberá, por tanto, sopesarse con el aumento de la potencia que se podría obtener al incluir más casos.

- Cuando el emparejamiento se utiliza como estrategia de muestreo, la decisión de emparejar se debe tomar al comienzo del estudio. Por lo tanto, es irreversible. Esto impide el análisis ulterior del efecto de las variables emparejadas sobre la variable de respuesta. También puede crear un error grave si la variable de emparejamiento no es una variable constitucional, como la edad o el sexo, sino un intermediario en la vía causal entre la variable predictiva y la variable de respuesta. Por ejemplo, si un investigador que desea investigar los efectos de la ingesta de alcohol sobre el riesgo de IM emparejara por la concentración sérica de lipoproteínas de alta densidad (HDL), podría pasar por alto cualquier efecto beneficioso del alcohol que estuviera mediado por el aumento de las HDL. Aunque se puede producir el mismo error con las estrategias de la fase de análisis, el emparejamiento introduce el error en el estudio, de tal forma que no se puede revertir; con las estrategias de la fase de análisis, el error se puede evitar alterando el análisis.

- El análisis correcto de los datos emparejados requiere técnicas analíticas especiales (análisis emparejados) que comparen cada participante solo con su pareja, y no con participantes que tienen niveles diferentes de las variables de confusión. Esto significa que no se pueden incluir los casos para los que no se puede encontrar una pareja. En el estudio de consumo de marihuana y tumores de células germinales, 39 de los 187 casos no presentaron ningún amigo que actuara como testigo (6). Los autores tuvieron que excluir de los análisis emparejados a estos 39 casos. El uso de técnicas analíticas sin emparejamiento con datos emparejados puede llevar a resultados incorrectos (en general, sesgados hacia la ausencia de efectos), porque se viola la suposición de que los grupos se muestrean de manera independiente.

- Un último inconveniente del emparejamiento es la posibilidad de emparejamiento excesivo, que sucede cuando la variable de emparejamiento está asociada con la variable predictiva, pero no es una variable de confusión, porque no está asociada a la de respuesta. El emparejamiento excesivo puede disminuir la potencia de un estudio de casos y testigos, porque el análisis emparejado desecha los conjuntos de caso y testigo con el mismo nivel de exposición (apéndice 8A-3). Por ejemplo, en el estudio de marihuana y tumores de células germinales, el uso de amigos como testigos puede haber reducido la potencia, al aumentar la concordancia de la exposición entre los casos y sus testigos emparejados: los amigos podrían tender a tener patrones similares de consumo de marihuana.

Estudios oportunistas

En ocasiones, hay oportunidades de controlar las variables de confusión en la fase de diseño, incluso sin medirlas; los denominamos diseños «oportunistas», porque utilizan oportunidades poco habituales para controlar las variables de confusión. Un ejemplo, útil cuando se estudian los efectos inmediatos de exposiciones a corto plazo, son los estudios cruzados (cap. 8): todas las posibles variables de confusión que son constantes a lo largo del tiempo (p. ej., edad, raza, sexo, clase social, factores genéticos) se controlan porque cada participante solo se compara consigo mismo en un período de tiempo diferente.

Otro diseño oportunista supone un experimento natural, en el que los participantes están expuestos o no están expuestos a un factor de riesgo particular durante un proceso que, en la práctica, actúa de manera aleatoria (7). Por ejemplo, Lofgren y cols. (8) estudiaron los efectos de la discontinuidad de la asistencia hospitalaria aprovechando el hecho de que los pacientes ingresados después de las cinco de la tarde en su centro se asignaban alternativamente a residentes más antiguos que, o seguían atendiendo a los pacientes, o los transferían a otro equipo a la mañana siguiente. Observaron que, en los pacientes que eran transferidos, se solicitaba un 38 % más de pruebas (p = 0,01) y tenían una estancia media 2 días mayor (p = 0,06) que los que permanecían con el mismo equipo. Igualmente, Bell y Redelmeier (9) estudiaron los efectos de la dotación de personal de enfermería, comparando las respuestas de pacientes con determinados diagnósticos que fueron ingresados en fin de semana con los que fueron ingresados a diario. Observaron una mayor mortalidad por tres afecciones que habían previsto que se vieran afectadas por una disminución de la dotación de personal los fines de semana, aunque sin ningún aumento de la mortalidad en los pacientes hospitalizados por otras enfermedades.

A medida que se han ido aclarando las diferencias genéticas en cuanto a la susceptibilidad a una exposición, ha surgido como opción una estrategia denominada aleatorización mendeliana (10). Esta estrategia funciona porque, en polimorfismos genéticos habituales, el alelo que una persona recibe está determinado al azar en las familias, y no está relacionado con la mayoría de las variables de confusión. Por ejemplo, algunos granjeros que rocían a las ovejas con insecticida (para matar garrapatas, piojos, etc.) tienen problemas de salud, como cefalea y astenia, que podrían deberse o no a su exposición laboral a los insecticidas. Los investigadores (11) aprovecharon un

polimorfismo del gen de la paraoxonasa 1 que produce enzimas con diferente capacidad para hidrolizar el insecticida organofosforado (diazinón-oxón) que se utiliza en los desinfectantes para ovejas. Encontraron que los granjeros expuestos con problemas de salud tenían más probabilidad de tener alelos asociados a una reducción de la actividad de la paraoxonasa 1 que granjeros con una exposición similar, pero asintomáticos. Este hallazgo ofreció un dato sólido de una relación causal entre la exposición al desinfectante para ovejas y los problemas de salud.

Los experimentos naturales y la aleatorización mendeliana son ejemplos de un método más general para mejorar la inferencia causal en los estudios de observación: el uso de variables instrumentales. Se trata de variables asociadas a la variable predictiva de interés, pero no asociadas independientemente a la variable de respuesta. Que alguien ingrese en fin de semana, por ejemplo, se asocia a los niveles de personal, pero se pensó que, por lo demás, no se asociaba al riesgo de mortalidad (para los diagnósticos estudiados), por lo que puede considerarse que el ingreso en fin de semana es una variable instrumental. Del mismo modo, la actividad de la enzima paraoxonasa 1 se asocia a una posible toxicidad por el desinfectante para las ovejas, pero no se asocia, por lo demás, al mal estado de salud. Otros ejemplos de variables instrumentales son la extracción de números de lotería para investigar efectos retardados sobre la mortalidad del servicio militar durante la época de la Guerra de Vietnam (12), y si la supervivencia prolongada tras un cáncer de riñón en estadio temprano depende de la distancia entre el domicilio del paciente y la consulta de un urólogo que realiza nefrectomías parciales o de otro que solamente realiza nefrectomías radicales (13).

Volver al principio

▪ ABORDAJE DE LAS VARIABLES DE CONFUSIÓN EN LA FASE DE ANÁLISIS

Las estrategias de la fase de diseño, especificación y emparejamiento precisan que se decida, al comienzo del estudio, qué variables son factores de confusión, y los investigadores no pueden estimar posteriormente los efectos de estas variables de confusión sobre una variable de respuesta. Por el contrario, las estrategias en la fase de análisis mantienen abiertas las opciones del investigador para que pueda cambiar de opinión sobre qué variables va a controlar en el momento del análisis.

A veces existen varias variables predictivas, y cada una de ellas puede actuar como variable de confusión de las demás. Por ejemplo, aunque beber café, fumar, ser varón y el tipo de personalidad se asocian al IM, también se asocian entre sí. El objetivo es determinar cuáles de esas variables predictivas están asociadas independientemente con el IM y cuáles están asociadas con el IM solo porque están asociadas a otros factores de riesgo (causales). En esta sección se comentan los métodos analíticos para evaluar la contribución independiente de variables predictivas y estudios de observación. En la tabla 9-5 [1] se resumen estos métodos.

Estratificación

Al igual que la especificación y el emparejamiento, la estratificación asegura que solo se comparen casos y testigos (o los participantes expuestos y no expuestos) con niveles similares de una posible variable de confusión. Conlleva la separación de los participantes en estratos (subgrupos) según el nivel de una posible variable de confusión, y el examen posterior de la relación entre la variable predictiva y la de respuesta por separado en cada grupo. En el apéndice 9A se ilustra la estratificación. Considerando a los fumadores y los no fumadores por separado («estratificando por el tabaquismo»), pueden eliminarse los efectos de confusión de esta variable.

El apéndice 9A también ilustra la modificación del efecto, en la que la estratificación muestra que la asociación entre variable predictiva y de respuesta varía con (es modificada por) el nivel de un tercer factor. La modificación del efecto introduce una complejidad adicional, porque la relación entre la variable predictiva y la variable de respuesta ya no se puede resumir con una única medida de asociación. Por casualidad solo, los cálculos de la asociación en diferentes estratos rara vez serán exactamente iguales, y los hallazgos únicamente indican modificación del efecto cuando las variables varían mucho. Una modificación del efecto clínicamente significativa es infrecuente, y antes de concluir que está presente, se debe evaluar su significación estadística y, especialmente si se han estudiado muchos subgrupos (lo que aumenta la probabilidad de que al menos uno sea estadísticamente significativo debido al azar), ver si se puede replicar en otra población. La verosimilitud biológica, o su ausencia, también puede contribuir a la interpretación. El tema de la modificación del efecto también surge en los análisis de subgrupos en ensayos clínicos (cap. 11), y en los metaanálisis cuando se está considerando la homogeneidad (similitud) de los estudios (cap. 13).

La estratificación tiene la ventaja de la flexibilidad: al realizar varios análisis estratificados, los investigadores pueden decidir qué variables parecen ser variables de confusión e ignorar el resto. Esto se puede realizar combinando el conocimiento sobre las probables direcciones de las relaciones causales con análisis para determinar si los resultados de los análisis estratificados difieren notablemente de los de análisis no estratificados (v. apéndice 9A). La estratificación también tiene la ventaja de que es reversible: no hay que hacer elecciones al principio del estudio que podrían lamentarse más tarde.

ESTRATEGIA	VENTAJAS	INCONVENIENTES

Estratificación	• Fácilmente comprensible
	• Flexible y reversible; puede escoger qué variables estratificar tras la recopilación de los datos
		• Número de estratos limitado por el tamaño de la muestra necesario en cada estrato
		  ◦ Pueden considerarse pocas variables
		  ◦ Pocos estratos por covariable dan un control incompleto de la confusión
		• Las variables importantes deben haberse medido

Ajuste estadístico	• Pueden controlarse al tiempo múltiples variables de confusión
	• La información en variables continuas puede usarse totalmente
	• Flexible y reversible
		• El modelo puede no encajar:
		  ◦ Control incompleto de la confusión (si el modelo no encaja en la relación variable de confusión-respuesta)
		  ◦ Cálculos inexactos de la solidez del efecto (si el modelo no encaja en la relación variable predictiva-de respuesta)
		• Los resultados pueden ser difíciles de entender (muchas personas no entienden el significado de un coeficiente de regresión)
		• Las covariables importantes deben haberse medido

Puntuaciones de predisposición	• Pueden controlarse al tiempo múltiples variables de confusión • La información de variables continuas puede usarse totalmente • Aumenta la posibilidad de control de la confusión cuando reciben el tratamiento más personas que las que logran la respuesta • Si se usa un análisis estratificado o emparejado, no es necesario que se acepten suposiciones sobre los modelos • Flexible e irreversible • La ausencia de superposición de las puntuaciones de predisposición puede resaltar los grupos en los que es difícil o imposible controlar la confusión	• Los resultados pueden ser difíciles de entender • Las covariables importantes deben haberse medido • Solo puede realizarse para participantes expuestos y no expuestos con puntuaciones de predisposición que se superponen, lo que permite reducir el tamaño de la muestra

El principal inconveniente del análisis estratificado es el limitado número de variables que pueden controlarse al mismo tiempo. Por ejemplo, en el estudio del café y el IM, la edad, el tipo de personalidad, la presión arterial sistólica, el colesterol sérico y el consumo de cigarrillos podrían ser variables de confusión. Para estratificar estas cinco variables, con solo tres estratos para cada una, ¡se necesitarían 3^5 (= 243) estratos! Con tantos estratos, habrá algunos sin casos ni testigos, y serán estratos que no podrán usarse.

Para mantener un número suficiente de participantes en cada estrato, a menudo se divide una variable en dos estratos. Cuando los estratos son demasiado amplios, sin embargo, puede que la variable de confusión no se controle de forma adecuada. Por ejemplo, si en el

estudio anterior se estratificó la edad usando solo dos estratos (p. ej., edad < 50 años y edad ≥ 50 años), sería posible que quedara alguna confusión residual si en cada uno de los estratos de edad los participantes que beben más café son mayores y, por tanto, tienen mayor riesgo de sufrir un IM.

Ajuste

Se dispone de varias técnicas estadísticas para ajustar las variables de confusión. Estas técnicas *modelan* la naturaleza de la asociación entre las variables para aislar los efectos de las variables predictivas y las de confusión. Por ejemplo, en un estudio de los efectos de la plumbemia sobre el CI de niños, se podría examinar la formación de los padres como una posible variable de confusión. El ajuste estadístico podría modelar la relación entre los años de estudios de los padres y el CI del niño como una línea recta, en la que cada año de formación de los padres se asocia a un aumento fijo del CI del niño. Los CI de niños con diferentes valores de plumbemia podrían ajustarse para eliminar el efecto de la formación de los padres usando el método descrito en el apéndice 9B.

Muchas veces un investigador quiere ajustar simultáneamente varios posibles factores de confusión, como edad, sexo, raza y nivel educativo. Esto precisa la utilización de técnicas de ajuste multivariadas, como regresión lineal multivariada y regresión logística, o el análisis de los riesgos proporcionales de Cox. Estas técnicas tienen otra ventaja: permiten el uso de toda la información en variables continuas. Por ejemplo, es fácil ajustar el nivel educativo de los padres en intervalos de 1 año, en lugar de estratificarlo en tan solo algunas categorías. Además, se pueden utilizar términos de interacción para modelar la modificación del efecto entre las variables.

Sin embargo, el ajuste multivariado tiene diversos inconvenientes. El más importante es que el modelo puede encajar. Los paquetes estadísticos informáticos han hecho que estos modelos sean tan accesibles que el investigador puede no pararse a considerar si su uso es adecuado para las variables predictiva y de respuesta del estudio [2] . Tomando el ejemplo del apéndice 9B, el investigador debe examinar si la relación entre los años de formación de los padres y el CI del niño es realmente lineal. Si el patrón es muy diferente (p. ej., la inclinación de la línea se vuelve más pendiente al aumentar la formación), los intentos de ajustar el CI a la formación de los padres usando un modelo lineal serán imperfectos, y el cálculo del efecto independiente de la plumbemia será incorrecto.

En segundo lugar, los estadísticos resultantes muchas veces son difíciles de comprender. Esto constituirá un problema si se utilizan transformaciones de las variables (p. ej., elevar al cuadrado la formación de los padres) o si se incluyen términos de interacción. Los investigadores deben pasar todo el tiempo que sea necesario con un estadístico (o hacer los cursos necesarios) para asegurarse de que puedan explicar el significado de los coeficientes y de otros estadísticos muy derivados que tengan previsto describir. Como precaución de seguridad, es una buena idea comenzar siempre con análisis estratificados sencillos e intentar comprender lo que está ocurriendo si otros análisis más complicados ofrecen resultados muy diferentes.

Puntuaciones de predisposición

Las puntuaciones de predisposición pueden ser particularmente útiles para los estudios de observación de la eficacia del tratamiento para controlar la confusión por indicación: el problema de que los pacientes para los que está indicado un tratamiento (y, por lo tanto, prescrito) tienen a menudo mayor riesgo o son diferentes en otros aspectos de los que no reciben el tratamiento. Recuerde que para ser una variable de confusión, una variable debe asociarse tanto con la variable predictiva como con la variable de respuesta. En lugar de ajustar todos los demás factores que predicen una *respuesta*, el uso de las puntuaciones de predisposición conlleva la creación de un modelo multivariado para predecir la recepción del *tratamiento*. A cada participante puede asignársele entonces una probabilidad prevista de tratamiento: una «puntuación de predisposición». Esta puntuación puede usarse como única variable de confusión en el análisis estratificado o multivariado.

Por otro lado, pueden emparejarse los participantes que han recibido el tratamiento con los que no lo han recibido por sus puntuaciones de predisposición, y compararse las respuestas entre parejas emparejadas. Al contrario del uso del emparejamiento como estrategia en la fase de diseño (muestreo), el emparejamiento por predisposición es similar a otras estrategias de la fase de análisis, porque es reversible. Sin embargo, los análisis de predisposición emparejados fracasan en pacientes a los que no se puede emparejar porque sus puntuaciones de predisposición son próximas a 0 o a 1. Aunque esto reduce el tamaño de la muestra, puede tener una ventaja, porque, en estos participantes a los que no se puede emparejar, el análisis de la puntuación de predisposición ha identificado la ausencia de comparabilidad entre grupos y la imposibilidad de controlar los factores de confusión, lo que podría no haber sido evidente con otros métodos de análisis multivariado.

EJEMPLO 9-1. Análisis de predisposición

Gum y cols. (14) estudiaron de forma prospectiva a 6 174 adultos consecutivos con ecocardiografía de esfuerzo; 2 310 de ellos (37 %) estaban tomando ácido acetilsalicílico y 276 fallecieron en el período de seguimiento de 3,1 años. En análisis sin ajustes no se asoció el consumo de ácido acetilsalicílico con la mortalidad (el 4,5 % en ambos grupos). Sin embargo, cuando se emparejaron 1 351 pacientes que habían recibido ácido acetilsalicílico con 1 351 pacientes con la misma predisposición a recibirlo, pero a los que no se administró, la

mortalidad fue un 47 % menor en los que fueron tratados (p = 0,002).

Los análisis que usan estas puntuaciones de predisposición tienen varias ventajas. El número de posibles variables de confusión que pueden modelarse como predictivas de una intervención habitualmente es mayor que el número de variables que se pueden introducir en el modelo como factores predictivos de una respuesta, porque el número de personas tratadas es generalmente mucho mayor que el número de los que presentan la respuesta (2 310 en comparación con 276 en el ejemplo 9-1). Otro motivo de que se puedan incluir más factores de confusión es que no hay riesgo de «ajustar en exceso» el modelo de predisposición: se pueden incluir términos de interacción, términos cuadráticos y variables indicadoras múltiples (15). Finalmente, los investigadores habitualmente tienen más confianza en identificar los determinantes del tratamiento que los determinantes de la respuesta, porque las decisiones terapéuticas las han tomado los médicos basándose en un número escaso de características del paciente.

Al igual que otras técnicas multivariadas, el uso de puntuaciones de predisposición sigue necesitando que se identifiquen y se midan las posibles variables de confusión. Una limitación de esta técnica es que no proporciona información sobre la relación entre cualquiera de las variables de confusión y la respuesta: el único resultado corresponde a la variable predictiva (habitualmente un tratamiento) que se introdujo en el modelo. Sin embargo, como es una estrategia de la fase de análisis, no evita la realización también de análisis multivariados más tradicionales, y suelen realizarse ambos tipos.

Volver al principio

▪ OTROS PROBLEMAS EN LA CUANTIFICACIÓN DE LOS EFECTOS CAUSALES

Condicionamiento a un efecto compartido

El sesgo que produce el condicionamiento a un efecto compartido es bastante complicado, y en ocasiones los libros de texto de iniciación lo pasan por alto, porque la mayoría de sus explicaciones utilizan diagramas y notaciones abstractos. Por el contrario, primero ofreceremos algunos ejemplos de cómo podría suceder, y después intentaremos explicar qué significa el nombre.

Considérese un estudio de personas que han perdido al menos 7 kg el año previo. Un investigador encuentra que las personas que han hecho dieta tienen menor riesgo de cáncer que las que no la han realizado. ¿Piensa usted que la dieta ha permitido prevenir el cáncer en estas personas?

Si se para a pensar, probablemente responda que no, porque el cáncer también produce pérdida de peso. Puede imaginar que si alguien pierde peso sin un motivo aparente es mucho más probable que tenga un cáncer que alguien que pierde peso mientras hace dieta. *En las personas que han perdido peso*, si la pérdida de peso no estuvo producida por la dieta, es más probable que haya estado producida por algo más ominoso. Los investigadores crearon una asociación inversa entre la dieta y el cáncer, condicionándolo (restringiendo la atención) a un efecto compartido (la pérdida de peso, que está producida tanto por la dieta como por el cáncer).

Se va a presentar otro ejemplo. En recién nacidos con bajo peso al nacimiento, aquellos cuyas madres fumaron durante el embarazo tienen menor mortalidad en el primer año que aquellos cuyas madres no fumaron (16). ¿Debemos animar a más mujeres a que fumen durante el embarazo? ¡Es evidente que no! El motivo de esta observación es que el tabaquismo produce bajo peso al nacimiento, pero también lo hacen otros factores, especialmente la prematuridad. Por lo tanto, *en recién nacidos con bajo peso al nacimiento*, si el bajo peso al nacimiento no estuvo producido por el tabaquismo, es más probable que haya estado producido por la prematuridad. Los investigadores crearon una asociación inversa entre el tabaquismo y la prematuridad (y su riesgo de mortalidad asociado), condicionando (restringiendo la atención) a un afecto compartido (bajo peso al nacimiento, que está producido tanto por el tabaquismo como por la prematuridad).

Ahora tiene sentido la frase «condicionamiento a un efecto compartido». Condicionamiento es un término epidemiológico que se refiere a la búsqueda de asociaciones entre las variables predictivas y las variables de respuesta «condicionada» (es decir, a niveles especificados) a algún atributo. Un efecto compartido se refiere a un atributo (como perder peso o ser un recién nacido con bajo peso al nacimiento) que tiene varias causas. Puede producirse sesgo por condicionamiento a un efecto compartido si el investigador trata algo *producido* por el factor de riesgo que se estudia como criterio de inclusión, variable de emparejamiento o posible variable de confusión.

Infravaloración de los efectos causales

Hasta este momento hemos puesto énfasis en la evaluación de la probabilidad de explicaciones alternativas para una asociación, a fin de evitar concluir que una asociación es real y causal cuando no lo es. Sin embargo, también es posible otro tipo de error: la *infraestimación* de los efectos causales. El azar, el sesgo y la confusión también pueden ser los motivos por los que se podría pasar por alto o infraestimar una asociación real.

Hemos hablado del azar como una razón para pasar por alto una asociación en el capítulo 5, al revisar los errores de tipo II y la necesidad de asegurar que el tamaño de la muestra proporcionará una potencia adecuada para hallar asociaciones reales. Sin embargo, tras haber completado un estudio, el cálculo de la potencia ya no es una buena forma de cuantificar la duda debida al error aleatorio. En esta fase, la potencia hipotética del estudio para detectar un efecto de una magnitud especificada es menos importante que los hallazgos reales expresados, como el cálculo observado de asociación (p. ej., cociente de riesgos) y su intervalo de confianza del 95 % (17).

El sesgo también puede alterar los cálculos de la asociación hacia la ausencia de efecto. En el capítulo 8, la necesidad del enmascaramiento para averiguar el estado de los factores de riesgo en los casos y los testigos era evitar el sesgo de medición diferencial, por ejemplo, las diferencias entre los casos y los testigos en la forma en la que se plantean las preguntas o se interpretan las respuestas, lo que podría llevar a los observadores a conseguir las respuestas que desean. Como los observadores podrían desear resultados en cualquier dirección, el sesgo de medida diferencial puede sesgar los resultados para que sobrestimen o para que infraestimen los efectos causales. Por otro lado, el sesgo no diferencial generalmente llevará a una infraestimación de las asociaciones.

La confusión puede también llevar a la atenuación de asociaciones reales. Por ejemplo, suponga que beber café realmente protege frente al IM, pero es más frecuente en los fumadores. Si no se controla el consumo de tabaco, los efectos beneficiosos del café podrían perderse: los consumidores de café aparecerían como pacientes con el mismo riesgo de IM que los que no toman café, cuando su mayor prevalencia de tabaquismo debería haber hecho que su riesgo fuera mayor. Este tipo de confusión, en la que se ocultan los efectos de un factor beneficioso por su asociación con una causa de la respuesta, se denomina a veces supresión (18). Es un problema habitual de los estudios de observación de los tratamientos, porque estos están indicados la mayoría de las veces en los pacientes que tienen mayor riesgo de una mala evolución. El resultado, señalado precozmente, es que puede parecer que un tratamiento beneficioso es inútil (como el ácido acetilsalicílico en el ejemplo 9-1) o incluso perjudicial, hasta que se controla la confusión por indicación.

Volver al principio

▪ ELECCIÓN DE UNA ESTRATEGIA

¿Qué orientaciones generales se pueden ofrecer para decidir si se deben abordar las variables de confusión durante las fases de diseño o análisis y cuál es la mejor forma de hacerlo? El uso de la especificación para controlar la confusión es el más adecuado en situaciones en las que el investigador está interesado principalmente en subgrupos específicos de la población; es realmente una forma especial del proceso general de establecer criterios para la selección de los participantes del estudio (cap. 3). Sin embargo, en estudios en los que el objetivo es la inferencia causal, se debe tener la precaución adicional de evitar criterios de inclusión que podrían estar producidos por las variables predictivas que se quieren estudiar (es decir, condicionamiento a un efecto compartido).

Una importante decisión que hay que tomar en la fase de diseño del estudio es si se debe realizar o no emparejamiento. Tiene su máxima utilidad en los estudios de casos y testigos y con factores constitucionales fijos, como la edad, la raza y el sexo. El emparejamiento también puede ser útil cuando el tamaño de la muestra es pequeño en comparación con el número de estratos necesarios para controlar variables de confusión conocidas, y cuando es más fácil emparejar las variables de confusión que medirlas. Sin embargo, como el emparejamiento puede comprometer permanentemente la capacidad del investigador de observar asociaciones reales, debe usarse poco, especialmente con variables que puedan estar en la cadena causal. En muchas situaciones, las estrategias de la fase de análisis (estratificación, ajuste y puntuaciones de predisposición) son igual de válidas para controlar la confusión, y tienen la mayor ventaja de ser reversibles: permiten que el investigador añada o quite covariables al modelo estadístico para explorar diferentes modelos causales.

Aunque no está disponible para todas las preguntas de la investigación, siempre merece la pena considerar la posibilidad de un diseño de estudio oportunista. Si no se para a considerar estos estudios (y a preguntar a sus compañeros), podría pasar por alto una gran oportunidad de hacer uno.

La decisión final para estratificar, ajustar o usar puntuaciones de predisposición se debe tomar cuando se hayan reunido los datos; en muchos casos, el investigador puede desear realizar todo lo anterior. Sin embargo, es importante, durante el diseño del estudio, que se considere qué factores pueden usarse más adelante para el ajuste, con el fin de saber qué variables medir. Además, como diferentes estrategias para controlar la confusión en la fase de análisis no siempre ofrecen los mismos resultados, es mejor especificar por adelantado un plan de análisis principal. Esto podría ayudar a los investigadores a resistir la tentación de seleccionar la estrategia que ofrezca los resultados más deseados.

Pruebas a favor de la causalidad

Hasta ahora, el método para aumentar la inferencia causal ha sido negativo: cómo descartar las cuatro explicaciones alternativas de la tabla 9-1 . Una estrategia complementaria consiste en buscar características de asociación que proporcionen pruebas positivas de causalidad, de las que las más importantes son la congruencia y la solidez de la asociación, la presencia de una relación de dosis-

respuesta y la verosimilitud biológica.

Cuando los resultados son congruentes en estudios de varios diseños, es menos probable que el azar o el sesgo sea la causa de una asociación. Sin embargo, también deben observarse las asociaciones reales que representan la relación de efecto y causa o de confusión. Por ejemplo, si los fumadores de cigarrillos beben más café y sufren más IM en la población, los estudios observarán de manera uniforme una asociación entre beber café y el IM.

La solidez de la asociación también es importante. Por un lado, las asociaciones sólidas proporcionan valores de *p* más significativos, con lo que el azar es una explicación menos probable. Las asociaciones sólidas también proporcionan mejores pruebas de causalidad, al reducir la probabilidad de confusión. Las asociaciones debidas a confusión son indirectas (a través de la variable de confusión) y, por lo tanto, suelen ser más débiles que las asociaciones directas de causa y efecto. Esto se muestra en el apéndice 9A: las asociaciones sólidas entre el café y el tabaco (razón de posibilidades = 16) y entre el tabaco y el IM (razón de posibilidades = 4) condujeron a una asociación mucho más débil entre el café y el IM (razón de posibilidades = 2,25).

Una relación de dosis-respuesta proporciona pruebas positivas de causalidad. La asociación entre fumar cigarrillos y el cáncer de pulmón es un ejemplo: los fumadores moderados tienen mayor incidencia de cáncer que los no fumadores, y los grandes fumadores presentan una incidencia incluso mayor. Siempre que sea posible, las variables predictivas deben medirse de forma continua o en varias categorías, de modo que pueda observarse cualquier relación de dosis-respuesta que exista. De nuevo, sin embargo, puede observarse una relación de este tipo con asociaciones de efecto y causa o con confusión.

Finalmente, la verosimilitud biológica es una consideración importante para extraer una inferencia causal: si puede proponerse un mecanismo causal que tenga sentido biológico, se refuerzan las pruebas para la causalidad, mientras que las asociaciones que no lo hacen, dado nuestro conocimiento actual de la biología, tienen menos probabilidad de representar una relación de causa y efecto. Por ejemplo, en el estudio del consumo de marihuana como factor de riesgo de tumores de células germinales, el consumo de marihuana menos de una vez al día se asoció a menor riesgo que la ausencia de consumo (6). Es difícil explicar este resultado desde la perspectiva biológica.

Es importante, sin embargo, no exagerar la verosimilitud biológica. Los investigadores parecen ser capaces de encontrar un mecanismo verosímil para prácticamente todas las asociaciones, y algunas asociaciones que originalmente se descartaron como no creíbles desde la perspectiva biológica, como la causa bacteriana de la enfermedad ulcerosa péptica, han resultado ser reales.

Volver al principio

▪ RESUMEN

- En el diseño de los estudios de observación se debe prever la necesidad de interpretar asociaciones. La inferencia de que la asociación representa una relación de causa y efecto (a menudo el objetivo del estudio) se refuerza mediante estrategias para reducir la probabilidad de las cuatro explicaciones alternativas: azar, sesgo, efecto y causa, y confusión.

- El papel del azar (error aleatorio) puede reducirse al mínimo diseñando un estudio con un tamaño de la muestra y una precisión adecuados para asegurar unas tasas bajas de errores de tipo I y II. Una vez completado el estudio, se puede juzgar el efecto del error aleatorio por la amplitud del intervalo de confianza del 95 % y la congruencia de los resultados con los datos previos.

- El sesgo (error sistemático) surge de las diferencias entre la población y los fenómenos que aborda la pregunta de la investigación, por un lado, y los participantes y medidas reales del estudio, por otro. Puede minimizarse basando las decisiones del diseño en el juicio de si estas diferencias llevarán a una respuesta equivocada a la pregunta de investigación.

- La relación de efecto y causa es menos probable si se diseña un estudio que permita la valoración de la secuencia temporal, y considerando la verosimilitud biológica.

- La confusión, que puede estar presente cuando una tercera variable está asociada con la variable predictiva de interés y es una causa de la variable de respuesta, es menos probable cuando se utilizan las siguientes estrategias, la mayoría de las cuales necesitan que se prevean y midan las posibles variables de confusión:

    - Especificación o emparejamiento en la fase de diseño, que alteran la estrategia de muestreo para asegurar que solo se comparan grupos con niveles similares de la variable de confusión. Estas estrategias deben usarse con prudencia porque pueden limitar irreversiblemente la información disponible del estudio.

    - Estrategias en la fase de análisis que consiguen el mismo objetivo y mantienen las opciones para investigar las vías

causales:

- Estratificación, que además de controlar el factor de confusión puede mostrar modificación del efecto («interacción»), una magnitud diferente de la asociación entre la variable predictiva y la variable de respuesta a diferentes niveles de una tercera variable.
- Ajuste, que puede permitir controlar simultáneamente el efecto de muchas variables predictivas.
- Puntuaciones de predisposición, que incrementan la potencia para abordar la confusión por indicación en estudios de observación de la eficacia del tratamiento.

- Los investigadores deben estar a la expectativa de diseños de observación oportunistas, como experimentos naturales, aleatorización mendeliana y otros diseños con variables instrumentales, que ofrecen una solidez de las inferencias causales que puede acercarse a la de un ensayo clínico aleatorizado.
- Los investigadores deben evitar el condicionamiento a efectos compartidos en la fase de diseño, no seleccionando a los participantes de acuerdo con covariables que podrían estar producidas por la variable predictiva, y en la fase de análisis no controlando esas covariables.
- Puede mejorarse la inferencia causal mediante pruebas positivas, sobre todo la congruencia y la solidez de la asociación, la presencia de una relación de dosis-respuesta y la verosimilitud biológica.

Volver al principio

[1]Surgen preguntas similares en estudios de pruebas diagnósticas (cap. 12), aunque, en esta situación, el objetivo no es determinar un efecto causal, sino determinar si la prueba en estudio aporta una capacidad predictiva importante a la información de que ya se disponía en el momento en que se realizó.

[2]Uno de nuestros compañeros bioestadísticos ha afirmado, en tono de broma, que intentar diseñar un paquete de programas estadísticos amigable e intuitivo es como intentar diseñar un coche de manera que un niño pueda llegar a los pedales.

[3]Esta descripción del análisis de la covarianza (ANCOVA) está simplificada. En realidad, los años de formación de los padres se representan en relación con el CI del niño en los grupos de plumbemia normal y elevada, y se utiliza la pendiente única que mejor ajusta los dos gráficos. Por lo tanto, el modelo para esta forma de ajuste asume que hay relaciones lineales entre la formación y el CI en ambos grupos, y que las pendientes de las líneas en los dos grupos son las mismas.

APÉNDICE 9AEjemplo hipotético de confusión y modificación del efecto
NA

Las entradas de estas tablas son números de participantes en este estudio de casos y testigos hipotético.

Panel 1. Si se analiza el grupo completo de participantes en el estudio, parece haber una asociación entre el consumo de café y el IM:

	Fumadores y no fumadores combinados	
	IM	SIN IM
Café	90	60
Sin café	60	90

Panel 2. Sin embargo, esto podría deberse a confusión, como se muestra en las tablas estratificadas por tabaquismo, en las que se ve que beber café no se asocia a IM ni en fumadores ni en no fumadores:

	Fumadores	
	IM	SIN IM
Café	80	40
Sin café	20	10

	No fumadores	
	IM	SIN IM
Café	10	20
Sin café	40	80

Razones de posibilidades para el IM asociado al café:

Fumar es una variable de confusión, porque está intensamente asociada a beber café (abajo, panel izquierdo) y al IM (abajo, panel derecho): estas tablas se obtuvieron reorganizando los números del panel 2.

	Con IM y sin IM combinado	
	Café	Sin café
Fumadores	120	30
No fumadores	30	120

	Con café y sin café combinados	
	Café	Sin café
Fumadores	100	50
No fumadores	50	100

Panel 3. La asociación entre el consumo de café y el IM del panel 1 también podría representar una modificación del efecto, si la estratificación por el tabaquismo revelara que la asociación entre el consumo de café y el IM difiere en fumadores y no fumadores. En la tabla siguiente, la RP de 2,25 para la asociación entre el consumo de café y el IM en fumadores y no fumadores combinados se debe por completo a una intensa asociación en fumadores. Cuando hay modificación del efecto, las razones de posibilidades en diferentes estratos son diferentes, y se deben describir por separado:

	Fumadores	
	IM	SIN IM
Café	50	15
Sin café	10	33

	No fumadores	
	IM	SIN IM
Café	40	45
Sin café	50	57

Razones de posibilidades de IM asociadas al café:

Comentario final: la asociación general entre el consumo de café y el IM del panel 1 podría estar ocultando la presencia de confusión por el tabaquismo, que se revelaría estratificando por el tabaquismo (panel 2). También podría estar ocultando la presencia de modificación del efecto por el tabaquismo, lo que también se observaría por la estratificación por el tabaquismo (panel 3). Asimismo, podría representar una relación de causa y efecto, que estaría respaldada (aunque no demostrada) si la estratificación por el tabaquismo no alterara la asociación entre el consumo de café y el IM. Finalmente (y con mayores visos de realismo), podría ser la consecuencia de alguna mezcla de todo lo anterior.

APÉNDICE 9BUn ejemplo simplificado de ajuste
NA

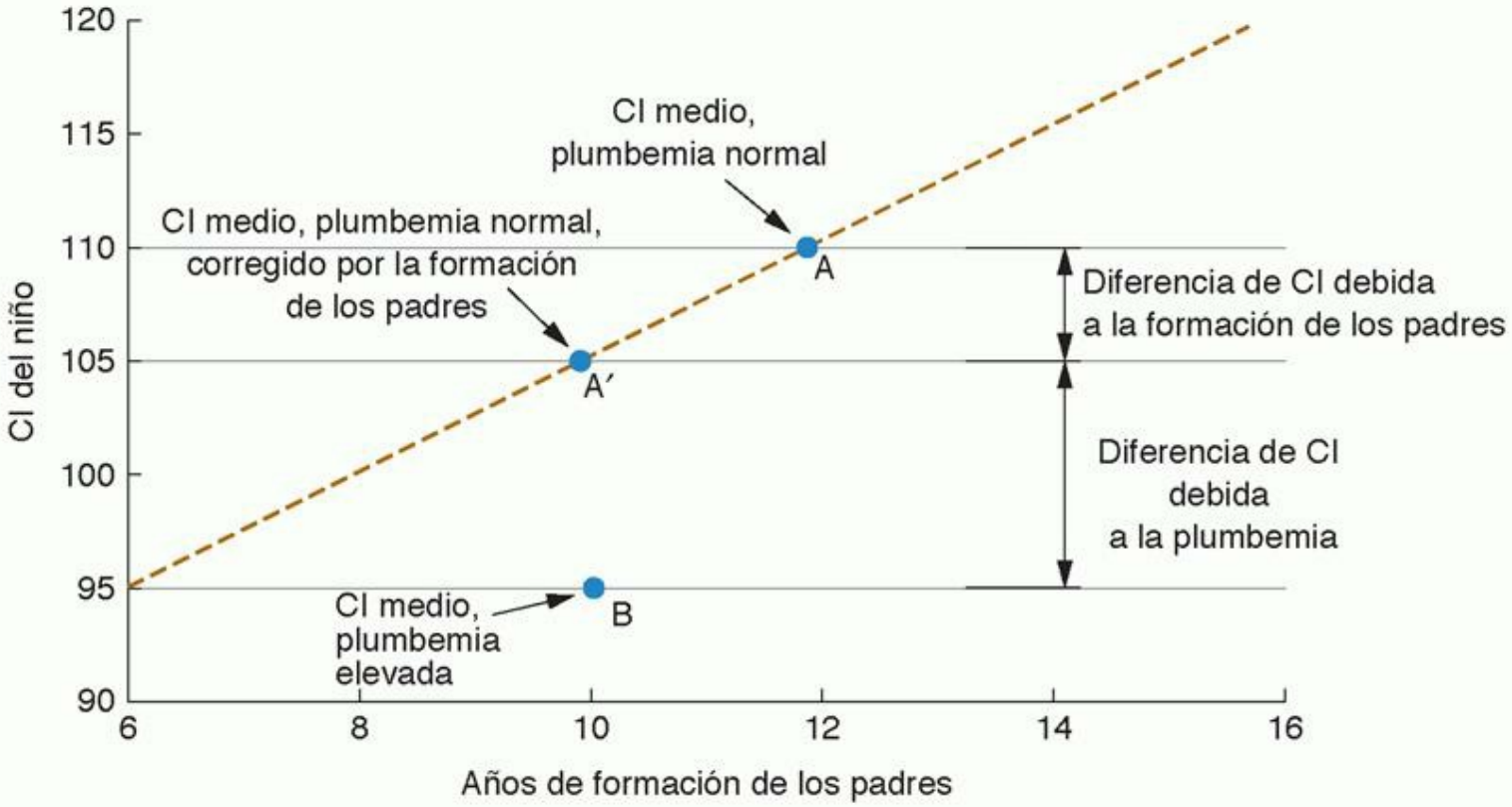
Suponga que en un estudio se observan dos importantes factores predictivos del cociente intelectual (CI) de los niños: el nivel de formación de los padres y la plumbemia del niño. Considere los siguientes datos hipotéticos en niños con plumbemias normales y elevadas:

	Promedio de años de formación de los padres	CI medio del niño
Plumbemia elevada	10,0	95

Plumbemia normal	12,0	110

Observe que el nivel de formación de los padres también se asocia con la plumbemia del niño. La pregunta es: «¿Es la diferencia de CI entre niños con plumbemia normal y elevada mayor de la que podría explicarse basándose en la diferencia de formación de los padres?» Para responder a esta pregunta, miramos qué diferencia de CI podría esperarse que se produjera por la diferencia de niveles de formación de los padres. Podemos hacerlo representando el nivel de formación de los padres frente al CI en niños con plumbemia normal (fig. 9-2) [3] .

La línea de puntos diagonal de la figura 9-2 muestra la relación entre el CI del niño y la formación de los padres en niños con plumbemia normal; existe un aumento del CI de 5 puntos por cada 2 años de formación de los padres. Por lo tanto, podemos ajustar el CI del grupo con plumbemia normal para explicar la diferencia en la formación media de los padres deslizando hacia abajo la línea desde el punto *A* hasta el punto *A′*. (Debido a que el grupo con plumbemia normal tiene 2 años más de formación de los padres en promedio, ajustamos sus CI hacia abajo en 5 puntos para que sean comparables con el grupo con plumbemia elevada en cuanto a la duración media de la formación de los padres.) Esto deja todavía una diferencia de 10 puntos en el CI entre los puntos *A* y *B*, lo que sugiere que el plomo tiene un efecto independiente de esta magnitud sobre el CI. Por lo tanto, de la diferencia de 15 puntos en el CI de niños con plumbemia baja y elevada, 5 pueden explicarse por los diferentes niveles de formación de sus padres, y los 10 restantes son atribuibles a la exposición al plomo.



▪FIGURA 9-2. Gráfica hipotética del CI del niño como una función lineal (línea discontinua) de los años de formación de los padres.
Herramientas de imágenes

BIBLIOGRAFÍA

1. McEvoy SP, Stevenson MR, McCartt AT, et al. Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. BMJ 2005;331(7514):428. Citado aquí

2. Magruder JT, Elahi D, Andersen DK. Diabetes and pancreatic cancer: chicken or egg? Pancreas 2011;40(3):339-351. Citado aquí

3. Huxley R, Ansary-Moghaddam A, Berrington de Gonzalez A, et al. Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies. Br J Cancer 2005;92(11):2076-2083. Citado aquí

4. Bosetti C, Rosato V, Polesel J, et al. Diabetes mellitus and cancer risk in a network of case-control studies. Nutr Cancer

2012;64(5):643-651. Citado aquí

5. Maconochie N, Doyle P, Carson C. Infertility among male UK veterans of the 1990-1 Gulf war: reproductive cohort study. BMJ 2004;329(7459):196-201. Citado aquí

6. Trabert B, Sigurdson AJ, Sweeney AM, et al. Marijuana use and testicular germ cell tumors. Cancer 2011;117(4): 848-853. Citado aquí

7. Newman TB, Kohn M. Evidence-based diagnosis. New York: Cambridge University Press, 2009. Chapter 10. Citado aquí

8. Lofgren RP, Gottlieb D, Williams RA, et al. Post-call transfer of resident responsibility: its effect on patient care [see comments]. J Gen Intern Med 1990;5(6):501-505. Citado aquí

9. Bell CM, Redelmeier DA. Mortality among patients admitted to hospitals on weekends as compared with weekdays. N Engl J Med 2001;345(9):663-668. Citado aquí

10. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003;32(1):1-22. Citado aquí

11. Cherry N, Mackness M, Durrington P, et al. Paraoxonase (PON1) polymorphisms in farmers attributing ill health to sheep dip. Lancet 2002;359(9308):763-764. Citado aquí

12. Hearst N, Newman TB, Hulley SB. Delayed effects of the military draft on mortality. A randomized natural experiment. N Engl J Med 1986;314(10):620-624. Citado aquí

13. Tan HJ, Norton EC, Ye Z, et al. Long-term survival following partial vs radical nephrectomy among older patients with early-stage kidney cancer. JAMA 2012;307(15):1629-1635. Citado aquí

14. Gum PA, Thamilarasan M, Watanabe J, et al. Aspirin use and all-cause mortality among patients being evaluated for known or suspected coronary artery disease: A propensity analysis. JAMA 2001;286(10):1187-1194. Citado aquí

15. Klungel OH, Martens EP, Psaty BM, et al. Methods to assess intended effects of drug treatment in observational studies are reviewed. J Clin Epidemiol 2004;57(12):1223-1231. Citado aquí

16. Hernandez-Diaz S, Schisterman EF, Hernan MA. The birth weight "paradox" uncovered? Am J Epidemiol 2006; 164(11):1115-1120. Citado aquí

17. Bacchetti P. Current sample size conventions: flaws, harms, and alternatives. BMC Med 2010;8:17. Citado aquí

18. Katz MH. Multivariable analysis: a practical guide for clinicians, 2nd ed. Cambridge, UK; New York: Cambridge University Press, 2006. Citado aquí

CAPÍTULO 10 Diseño de un estudio aleatorizado con enmascaramiento

Steven R. Cummings
Deborah Grady
Stephen B. Hulley

En los ensayos clínicos, el investigador aplica una intervención y observa el efecto sobre una o más variables de respuesta. La principal ventaja de un ensayo sobre un estudio de observación es la capacidad de demostrar la causalidad. La asignación aleatoria de la intervención minimiza la influencia de las variables de confusión, y el enmascaramiento de su administración minimiza la posibilidad de que los efectos aparentes de la intervención se deban al uso diferencial de otros tratamientos en los grupos de intervención y testigo, o a la determinación y la adjudicación sesgadas del resultado.

Sin embargo, los ensayos clínicos suelen ser caros y prolongados, abordan preguntas limitadas y a veces exponen a los participantes a posibles peligros. Por estos motivos se reservan para preguntas de investigación relativamente maduras, en las que los estudios de observación y otros tipos de datos sugieren que una intervención podría ser eficaz y segura, pero se precisan pruebas más claras antes de que se pueda aprobar o recomendar. No todas las preguntas de investigación pueden tratarse mediante un diseño de ensayo clínico: no puede estudiarse si el tratamiento farmacológico del aumento del colesterol-LDL en niños evitará los infartos de miocardio muchas décadas después, y no es ético aleatorizar a las personas a fumar cigarrillos reales o simulados para determinar el efecto sobre el cáncer de pulmón. Sin embargo, siempre que sea posible, se deben obtener datos de estudios clínicos sobre la eficacia y la seguridad de las intervenciones clínicas.

Este capítulo se centra en el diseño del clásico ensayo aleatorizado con enmascaramiento: seleccionando las condiciones de los grupos de intervención y testigo, definiendo las variables de respuesta y los efectos adversos, seleccionando los participantes, determinando las variables predictivas y las variables de respuesta, y evaluando los abordajes de la aleatorización y el enmascaramiento. En el capítulo siguiente se expondrán otros diseños de ensayos clínicos y problemas relacionados con su ejecución y análisis.

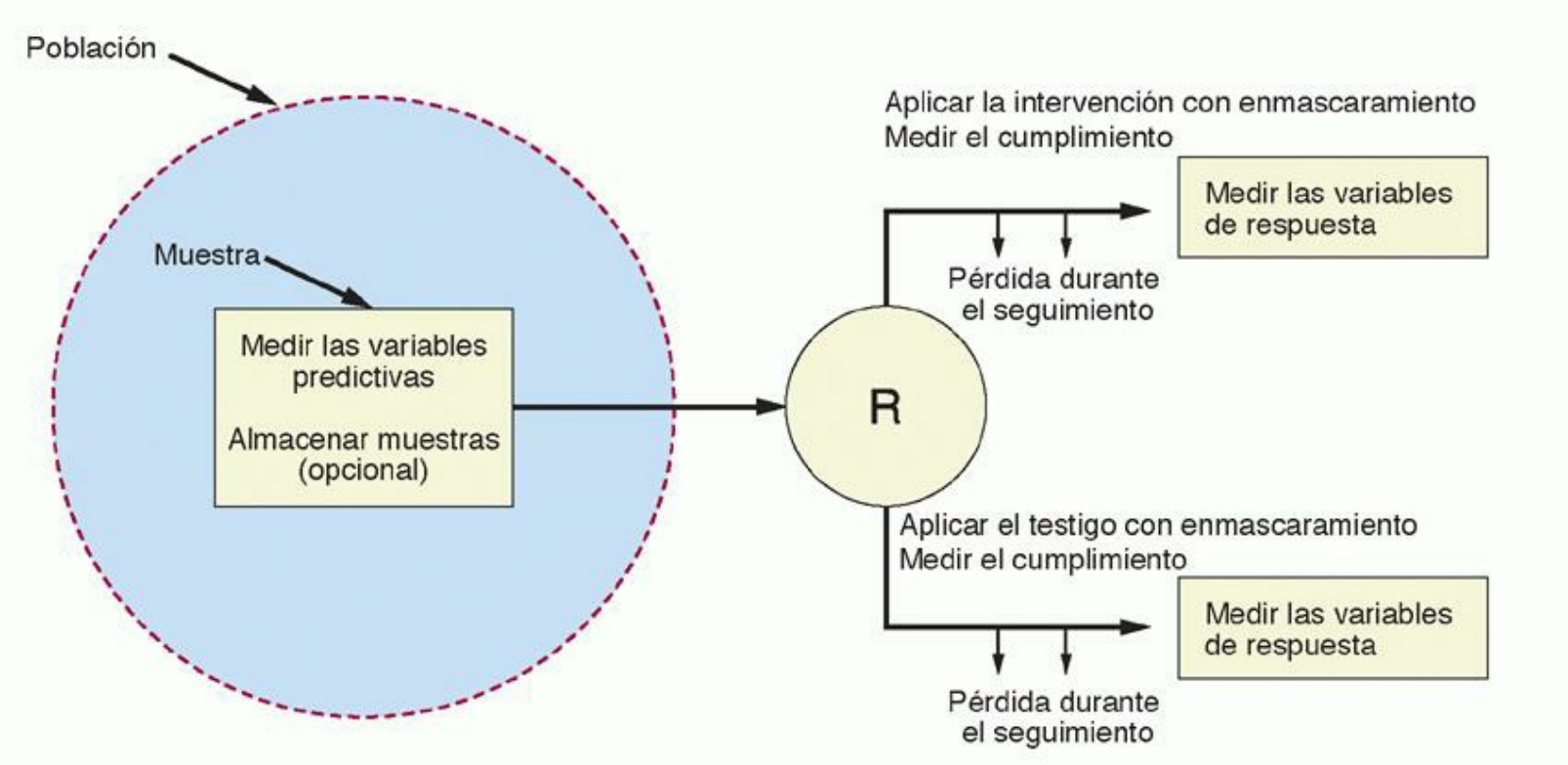
▪ SELECCIÓN DE LAS CONDICIONES DE INTERVENCIÓN Y TESTIGO

El ensayo aleatorizado clásico es un diseño de grupos paralelos, que incluye un grupo que recibe una intervención que se quiere estudiar y un grupo testigo que no recibe ningún tratamiento activo (preferiblemente un placebo) o recibe un tratamiento de comparación. El investigador aplica la intervención y el tratamiento testigo, sigue a los dos grupos a lo largo del tiempo y compara la variable de respuesta entre los grupos de intervención y testigo (fig. 10-1).

Elección de la intervención

La elección de la intervención es el primer paso esencial en el diseño de un ensayo clínico. Los investigadores deben considerar varios puntos a la hora de diseñar las intervenciones, como la dosis, la duración y la frecuencia de la intervención que mejor equilibre eficacia y seguridad. También es importante tener en cuenta la viabilidad del enmascaramiento, si se va a tratar con una intervención o con una combinación de intervenciones, la aceptabilidad para los participantes y la posibilidad de generalizar la forma en la que el tratamiento se usará en la práctica. Si hay incertidumbre sobre decisiones importantes, como cuál es la dosis que equilibra mejor eficacia y seguridad, es mejor, generalmente, posponer los ensayos importantes o costosos hasta que se hayan realizado estudios preliminares que ayuden a resolver el tema.

El mejor equilibrio entre eficacia y seguridad depende de la intervención y la afección que se esté estudiando. Por un lado, la eficacia suele ser la consideración principal en el diseño de intervenciones para tratar enfermedades que causan síntomas graves o el fallecimiento. Por lo tanto, puede que sea mejor escoger la mayor dosis tolerable para el tratamiento del cáncer metastásico. Por otro lado, la seguridad debe ser el principal criterio para diseñar intervenciones para tratar enfermedades sintomáticas que raras veces llevan a la progresión de la enfermedad o la muerte. El tratamiento preventivo en personas sanas debe cumplir estrictas pruebas de seguridad: si es eficaz, el tratamiento evitará la afección en algunas personas, pero todo aquel al que se haya tratado correrá el riesgo de sufrir los efectos secundarios del tratamiento. En este caso, suele ser mejor elegir la dosis que aumente al máximo la eficacia con un riesgo muy bajo de efectos adversos. Si la mejor dosis no es segura según los hallazgos previos de investigación en animales y en seres humanos, puede que sea preciso realizar ensayos adicionales que comparen los efectos de múltiples dosis sobre marcadores intermedios o variables de respuesta clínicas (v. ensayos de fase III, cap. 11).



▪FIGURA 10-1. En un estudio aleatorizado con enmascaramiento, los pasos son:
• Seleccionar una muestra de pacientes de una población adecuada para recibir la intervención.
• Medir las variables predictivas y, cuando proceda, los niveles iniciales de la variable de respuesta.
• Considerar la opción de almacenar suero, imágenes y otro tipo de datos para su análisis posterior.
• Asignar aleatoriamente y con enmascaramiento la intervención y la situación de testigo (p. ej., placebo).
• Seguir a la cohorte a lo largo del tiempo, minimizando las pérdidas durante el seguimiento y evaluando el cumplimiento de la intervención y del seguimiento.
• Medir las variables de respuesta.
Herramientas de imágenes

A veces, un investigador puede decidir comparar varias dosis o niveles de intensidad con un solo grupo testigo. Por ejemplo, cuando se diseñó el *Multiple Outcomes of Raloxifene Evaluation Trial* (estudio de evaluación de múltiples respuestas del raloxifeno), no estaba claro cuál era la mejor dosis de raloxifeno (60 o 120 mg), de modo que se probaron dos dosis para evitar las fracturas vertebrales (1). Esta es, a veces, una estrategia razonable, pero tiene su coste: un ensayo mayor y más caro, y la complejidad de tratar con múltiples hipótesis (cap. 5).

Con algunos tratamientos se ajusta la dosis para optimizar el efecto en cada paciente individual. En estos casos puede ser mejor diseñar una intervención, de forma que la dosis del fármaco activo se ajuste para lograr una respuesta clínica, como la reducción de la viremia en la hepatitis C. Para conservar el enmascaramiento deben realizarse los cambios correspondientes (por alguien que no intervenga en el ensayo) en la «dosis» del placebo para un participante seleccionado al azar o emparejado en el grupo tratado con placebo.

Los ensayos para estudiar intervenciones únicas suelen ser mucho más fáciles de planificar y realizar que los que estudian combinaciones de tratamientos. Sin embargo, muchas afecciones médicas, como la infección por el VIH y la insuficiencia cardíaca congestiva, se tratan con combinaciones de fármacos o tratamientos. El principal inconveniente del estudio de combinaciones de tratamientos es que el resultado no puede proporcionar conclusiones claras sobre cada uno de los elementos de las intervenciones. En el primer ensayo *Women's Health Initiative,* por ejemplo, se trató a mujeres posmenopáusicas con estrógenos y progestágenos o con placebo. La intervención aumentaba el riesgo de sufrir varias afecciones, como cáncer de mama; sin embargo, no estaba claro si el efecto se debía al estrógeno o al progestágeno (2). En general, es preferible diseñar ensayos clínicos que solo tengan una diferencia importante entre los dos grupos de estudio.

El investigador debe tener en cuenta en qué medida los participantes serán receptivos a la intervención propuesta y si se puede realizar con enmascaramiento. Otra consideración es la facilidad con la que se puede incorporar la intervención a la práctica habitual. Las intervenciones sencillas suelen ser generalmente mejores que las complicadas (es más probable que los pacientes se tomen una pastilla una vez al día que se administren inyecciones subcutáneas dos o tres veces al día). Puede no ser posible aplicar intervenciones complicadas con aspectos cualitativos, como el asesoramiento múltiple acerca del cambio de conducta, a la práctica general, porque son difíciles de replicar, laboriosas y costosas. Estas intervenciones tienen menos probabilidades de tener consecuencias sobre la salud

pública, incluso aunque un ensayo demuestre que son eficaces.

Elección de los testigos

El mejor grupo testigo recibe un tratamiento no activo de un modo que puede enmascararse, lo que, para medicaciones, generalmente precisa un placebo que no pueda distinguirse del tratamiento activo. Esta estrategia compensa cualquier efecto placebo de la intervención activa (por sugestión o expectativa), de modo que cualquier diferencia de respuesta entre los grupos del estudio pueda atribuirse a un efecto específico de la intervención.

La comparación más clara entre los grupos de intervención y testigo se produce cuando no existen cointervenciones: medicamentos, tratamientos o conductas (aparte de la intervención del estudio) que alteran el riesgo de producir el resultado de interés. Por ejemplo, en un ensayo clínico en el que se compara una intervención de yoga con el tratamiento habitual para prevenir la diabetes, el personal del estudio puede animar a los participantes a que realicen ejercicio y pierdan peso. Se trata de cointervenciones posiblemente eficaces que pueden reducir el riesgo de presentar diabetes. Si los participantes de los dos grupos utilizan cointervenciones eficaces, disminuirá la incidencia de la variable de respuesta, se reducirá la potencia, y el tamaño de la muestra deberá ser mayor o el estudio deberá ser más prolongado. Si el uso de cointervenciones eficaces difiere entre los grupos de intervención y testigo, el resultado estará sesgado. Si no existe un enmascaramiento eficaz, el protocolo debe incluir planes para obtener datos que permitan el ajuste estadístico de las diferencias entre los grupos en cuanto a la tasa de utilización de estas cointervenciones durante el estudio. No obstante, puede ser difícil medir las cointervenciones, y debe considerarse que el ajuste de estas diferencias posteriores a la aleatorización es un análisis secundario o explicativo, porque viola el principio de intención de tratar (cap. 11).

A menudo no es posible suspender otros tratamientos que no son la intervención del estudio. Por ejemplo, en un ensayo clínico de un nuevo fármaco para reducir el riesgo de sufrir un infarto de miocardio en personas con cardiopatía isquémica (CI) diagnosticada, los investigadores no pueden, desde el punto de vista ético, prohibir ni aconsejar a los participantes que no tomen tratamientos médicos que están indicados en pacientes con CI diagnosticada, como ácido acetilsalicílico, estatinas y β-bloqueantes. Una solución sería administrar el tratamiento farmacológico habitual a todos los participantes del ensayo; aunque este método reduce la incidencia general de episodios y, por lo tanto, aumenta el tamaño de la muestra necesario, permite estudiar la pregunta clínica más importante: si la nueva intervención mejora el resultado cuando se administra además del tratamiento habitual.

Cuando el tratamiento que se va a estudiar es un fármaco nuevo que probablemente sea una buena alternativa al tratamiento habitual, una opción será diseñar un ensayo de ausencia de inferioridad o de equivalencia en el que se compare el nuevo tratamiento con el que ya se ha demostrado que es eficaz (cap. 11).

Volver al principio

▪ ELECCIÓN DE LAS VARIABLES DE RESPUESTA

La definición de las variables de respuesta específicas del ensayo clínico influye en otros muchos componentes del diseño, así como en el coste y la viabilidad del estudio. Los estudios deben incluir, habitualmente, diversas variables de respuesta para aumentar la riqueza de los resultados y la posibilidad de realizar análisis secundarios. Sin embargo, se debe designar una de ellas que sea la variable de respuesta principal, que refleje la pregunta principal, permita el cálculo del tamaño de la muestra y establezca la prioridad de los esfuerzos para realizar el estudio.

Las variables de respuesta clínicas constituyen la mejor prueba sobre si se deben utilizar o no los tratamientos y las intervenciones preventivas, y cómo hacerlo. Para las respuestas que no son habituales, como la aparición de cáncer, los ensayos deben ser generalmente extensos, prolongados y caros. Como se indicó en el capítulo 6, las variables de respuesta medidas como variables continuas, como la calidad de vida, suelen poder estudiarse con menos participantes que las variables de respuesta dicotómicas. Sin embargo, en ocasiones, la variable de respuesta clínica más importante es, inevitablemente, dicotómica, como la recurrencia del cáncer.

Los marcadores intermedios, como la densidad ósea, son determinaciones que están relacionadas con la variable de respuesta clínica. Los ensayos que utilizan variables de respuesta intermedias pueden mejorar el conocimiento de la fisiopatología y proporcionar información para elegir la mejor dosis o frecuencia de tratamiento en los estudios con variables de respuesta clínicas. La importancia clínica de los ensayos con variables de respuesta intermedias depende de la exactitud con que los cambios de estos marcadores, en especial los cambios que se producen debido al tratamiento, representan cambios del riesgo de la variable de respuesta clínica. Puede considerarse que los marcadores intermedios son marcadores indirectos de la respuesta clínica en la medida en que los cambios inducidos por el tratamiento sobre el marcador permiten predecir de forma constante el modo en que el tratamiento cambia la variable de respuesta clínica (3). En general, un buen marcador indirecto mide los cambios en un factor intermedio de la vía principal que determina la respuesta clínica.

La viremia del VIH es un buen marcador indirecto, porque los tratamientos que reducen la viremia reducen también la morbilidad y la mortalidad de los pacientes con infección por el VIH. Por el contrario, se considera que la densidad mineral ósea (DMO) es mal marcador indirecto (3). Refleja la cantidad de mineral en una sección de un hueso, pero los tratamientos que mejoran la DMO a veces tienen un efecto escaso o nulo sobre el riesgo de fracturas, y la magnitud del cambio de la DMO no se relaciona de manera constante con la magnitud en que el tratamiento reduce el riesgo de sufrir fracturas (4). La mejor prueba de que un marcador biológico es un buen marcador indirecto procede de ensayos clínicos aleatorizados de la respuesta clínica (fracturas), que también miden el cambio del marcador (DMO) en todos los participantes. Si el marcador es un buen marcador indirecto, el ajuste estadístico de los cambios del marcador explicará gran parte del efecto del tratamiento sobre la variable de respuesta (3).

Número de variables de respuesta

A menudo es deseable tener varias variables de respuesta que midan aspectos diferentes de los fenómenos de interés. En el estudio *Heart and Estrogen/Progestin Replacement Study* (HERS) se eligieron los episodios de cardiopatía isquémica como criterio de valoración principal. También se evaluó el infarto de miocardio no mortal, la revascularización coronaria, la hospitalización por angina inestable o insuficiencia cardíaca congestiva, el accidente cerebrovascular, el accidente isquémico transitorio, los episodios de tromboembolia venosa y la mortalidad por todas las causas, y se realizó la adjudicación para obtener una descripción más detallada de los efectos cardiovasculares del tratamiento hormonal (5). Sin embargo, se designó una única variable de respuesta principal (episodios de CI) con el fin de planificar el tamaño de la muestra y la duración del estudio, además de evitar los problemas de interpretación de las pruebas de hipótesis múltiples (cap. 5).

Criterios de valoración compuestos

En algunos estudios se definen criterios de valoración formados por diferentes episodios o medidas diferentes. Por ejemplo, en muchos estudios de intervenciones para reducir el riesgo de cardiopatía isquémica se incluyen en el criterio de valoración varios episodios coronarios específicos, como infarto de miocardio, muerte coronaria y técnicas de revascularización coronaria. Esto es razonable si cada uno de estos resultados tiene importancia clínica, si el tratamiento actúa mediante mecanismos similares en cada una de las enfermedades y si se espera que la intervención reduzca el riesgo de todos esos episodios. Además, un criterio de valoración compuesto generalmente ofrece más potencia que un único criterio de valoración, porque habrá más episodios. Sin embargo, los criterios de valoración compuestos que incluyen episodios que no tienen tanto significado clínico o que se producen con mucha más frecuencia que los otros del criterio de valoración compuesto pueden llevar a hallazgos erróneos. Por ejemplo, si se añade el ingreso hospitalario para evaluación de dolor torácico al criterio de valoración coronario compuesto, este episodio dominará el criterio de valoración compuesto si estos ingresos hospitalarios se producen con mucha más frecuencia que el infarto de miocardio, la muerte coronaria y la revascularización. En este caso, se puede describir que una intervención que altera el criterio de valoración compuesto reduce el riesgo de «episodios coronarios», cuando en realidad solo reduce el riesgo de ingreso hospitalario por dolor torácico.

Los criterios de valoración compuestos se deben diseñar con cuidado. Si el tratamiento produce tan solo un efecto pequeño sobre un criterio de valoración, especialmente si ese criterio de valoración es relativamente frecuente, puede añadir poca potencia estadística o incluso puede aumentar el tamaño de la muestra necesario para detectar un efecto. Por ejemplo, si se añade el accidente cerebrovascular a un «criterio de valoración cardiovascular» compuesto, la intervención podría reducir el riesgo de episodios coronarios, pero podría no tener ninguna influencia o incluso podría aumentar el riesgo de accidente cerebrovascular, por lo que se encontraría que no tiene ningún efecto sobre el criterio de valoración cardiovascular compuesto.

Efectos adversos

El investigador debe incluir medidas de respuesta que detecten la aparición de efectos adversos que pueden deberse a la intervención. Dar a conocer si los efectos beneficiosos de una intervención valen más que los adversos es un objetivo principal de la mayor parte de los ensayos clínicos, incluso los que prueban tratamientos aparentemente inocuos, como un programa de educación sanitaria. Los efectos adversos pueden oscilar desde síntomas relativamente leves, como un exantema transitorio o leve, hasta complicaciones graves y mortales. La incidencia de aparición, el efecto del tratamiento y los requisitos de tamaño de la muestra para detectar efectos adversos generalmente son diferentes de los necesarios para detectar efectos beneficiosos. Desgraciadamente, los efectos secundarios no habituales serán, generalmente, imposibles de detectar incluso en estudios extensos, y solo se descubren (si lo hacen) con estudios de observación extensos o con series de casos después de que una intervención tenga un amplio uso clínico.

En las primeras etapas del estudio de un nuevo tratamiento, cuando los posibles efectos adversos no están claros, los investigadores deben realizar preguntas amplias, abiertas, sobre todos los tipos de posibles efectos adversos. En ensayos de gran tamaño, la evaluación y la codificación de todos los posibles efectos adversos pueden resultar muy caras y laboriosas, y a menudo producen pocos resultados importantes. Los investigadores deben considerar estrategias para reducir al mínimo esta carga mientras conservan una evaluación adecuada de los posibles daños debidos a la intervención. Por ejemplo, en ensayos de gran tamaño, los acontecimientos leves y habituales, como infecciones de las vías respiratorias superiores y molestias gastrointestinales, podrían registrarse en un subgrupo de los

participantes. Puede no ser necesario registrar los efectos adversos que no sean graves si en estudios previos no se han encontrado diferencias en la incidencia de síntomas leves. Además de estas preguntas abiertas, se deben diseñar cuestionarios específicos para descubrir acontecimientos adversos importantes que se esperan por la investigación o la experiencia clínica previa. Por ejemplo, debido a que la miositis es un efecto secundario comunicado del tratamiento con estatinas, deben investigarse los síntomas y signos de miositis en cualquier ensayo de una nueva estatina.

Los efectos adversos que se describen como síntomas o términos clínicos se deben codificar y categorizar para su análisis. Los diccionarios MedDRA (www.ich.org/products/meddra.html) y SNOMED (https://www.nlm.nih.gov/research/umls/) son diccionarios de términos utilizados habitualmente que están agrupados de diferentes maneras, como síntomas, diagnósticos específicos y sistema orgánico al que afectan. Por ejemplo, un acontecimiento adverso registrado como «fiebre y tos» y un acontecimiento adverso registrado como «bronquitis» se agruparán con otras enfermedades, como la neumonía, como una «infección respiratoria» y, en un nivel superior, como efectos adversos en el aparato respiratorio. Estos esquemas de clasificación ofrecen un buen resumen general de los acontecimientos adversos y tienen una exactitud razonable para enfermedades que se diagnostican específicamente, como las fracturas. Sin embargo, pueden pasar por alto acontecimientos adversos importantes que se describen con varios términos si dichos términos no se agrupan en conjunto. Por ejemplo, en un estudio de denosumab para la prevención de las fracturas osteoporóticas, con el diccionario MedDRA se codificaron los casos de celulitis por separado de los casos de erisipela (dos nombres para el mismo tipo de infección). Cuando se combinaron, se habían producido 12 casos graves de celulitis con denosumab, en comparación con 1 con el placebo ($p <$ 0,001) (6).

Los acontecimientos adversos también se clasifican generalmente por su gravedad. Se definen los acontecimientos adversos graves (AAG) como los episodios que producen la muerte o son potencialmente mortales, o los episodios que precisan o prolongan un ingreso hospitalario, o que producen discapacidad o daños permanentes, malformaciones congénitas u otros episodios médicos importantes que pueden precisar una intervención médica o quirúrgica para prevenir una de las demás consecuencias (www.fda.gov/Safety/MedWatch/HowToReport/ucm053 087.htm). Los acontecimientos adversos graves generalmente se deben notificar rápidamente a los comités de revisión institucionales y al promotor del estudio.

Cuando los datos de un ensayo se usan para solicitar la autorización de un nuevo fármaco por las autoridades de registro sanitario, el diseño del ensayo debe satisfacer las normas de las agencias de registro sanitario para la notificación de acontecimientos adversos (http://www.fda.gov/Drugs/InformationOnDrugs/ucm135 151.htm). Determinadas áreas de enfermedades, como el cáncer, han establecido métodos para clasificar los acontecimientos adversos (http://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm).

Volver al principio

▪ SELECCIÓN DE LOS PARTICIPANTES

En el capítulo 3 se expone cómo especificar los criterios de participación que definen una población objetivo que sea adecuada para la pregunta de la investigación y una población accesible cuyo estudio sea práctico, cómo diseñar un enfoque científico y eficaz para seleccionar participantes, y cómo reunirlos. Aquí se tratarán puntos especialmente relevantes para los ensayos clínicos.

Definir los criterios de participación

En un ensayo clínico, los criterios de inclusión y exclusión tienen el objetivo de identificar una población en la que es posible, ético y relevante estudiar el efecto de la intervención sobre las respuestas. Los criterios de inclusión deben producir un número suficiente de participantes que tengan una incidencia suficientemente elevada del criterio de valoración principal para lograr la potencia adecuada para encontrar un efecto importante de la intervención sobre el criterio de valoración. Por otro lado, los criterios también deben aumentar al máximo la posibilidad de generalizar los hallazgos del ensayo y facilitar la reunión de los pacientes. Por ejemplo, si la respuesta de interés es un acontecimiento infrecuente, como el cáncer de mama, suele ser necesario reunir participantes de riesgo elevado para reducir el tamaño de la muestra y el tiempo de seguimiento hasta unos niveles factibles. Por otro lado, si se limitan los criterios de inclusión a mujeres de riesgo elevado, se limita también la posibilidad de generalizar los resultados y se dificulta la inclusión de suficientes participantes en el ensayo.

Para planificar el tamaño de la muestra correcto, el investigador debe contar con estimaciones fiables de la incidencia del criterio de valoración principal en personas que podrían participar. Estos cálculos pueden basarse en datos de estadísticas vitales, estudios de observación longitudinales o incidencias observadas en el grupo sin tratamiento de ensayos clínicos con criterios de inclusión similares a los del estudio previsto. Por ejemplo, la incidencia esperada de cáncer pancreático en adultos se puede estimar a partir de datos de registros de cáncer. El investigador debe tener en cuenta, no obstante, que los efectos del cribado y del sesgo de voluntario sano generalmente indican que incluso las tasas de incidencia en los pacientes que cumplen los criterios de un estudio clínico y dan su

conformidad a participar en él son inferiores a las de la población general; puede ser preferible obtener las tasas de incidencia de cáncer pancreático del grupo tratado con placebo en otros ensayos con criterios de inclusión similares.

La inclusión de personas con un riesgo elevado de presentar la variable de respuesta puede reducir el número de participantes necesarios para el estudio. Si se han establecido los factores de riesgo de la respuesta, pueden diseñarse los criterios de selección para incluir participantes que tienen un mínimo riesgo estimado de la respuesta de interés. En el estudio *The Raloxifene Use for The Heart*, diseñado para comprobar el efecto del raloxifeno sobre la prevención de la enfermedad cardiovascular (ECV) y el cáncer de mama, se incluyó a mujeres con mayor riesgo de ECV según una combinación de factores de riesgo (7). Otro modo de aumentar la incidencia de los acontecimientos es limitar la inclusión a personas que ya sufren la enfermedad. En el estudio *The Heart and Estrogen/Progestin Replacement Study* se incluyó a 2 763 mujeres que ya sufrían enfermedad coronaria para comprobar si la combinación de estrógenos y progestágenos reducía el riesgo de sufrir nuevos episodios de CI (5). Este método era mucho más barato que el del estudio *Women's Health Initiative*, con la misma pregunta de investigación en mujeres sin CI, en el que se precisaron unas 17 000 participantes (8).

Aunque las muestras probabilísticas de la población general tienen ventajas en los estudios de observación, este tipo de muestreo no suele ser posible ni necesario en los ensayos aleatorizados. La inclusión de participantes con diversas características aumentará la confianza de que los resultados de un ensayo se podrán aplicar ampliamente. Sin embargo, salvo que haya entre las poblaciones diferencias biológicas y genéticas que influyan en el efecto de un tratamiento, suele ser cierto que los resultados de un ensayo clínico realizado en una muestra de conveniencia (p. ej., mujeres con CI que responden a un anuncio) serán similares a los resultados obtenidos en muestras de probabilidad de personas elegibles (todas las mujeres con coronariopatía). En ocasiones, la eficacia de un tratamiento depende de las características de los pacientes. Esto se denomina modificación del efecto o interacción (cap. 11). Por ejemplo, algunos tratamientos de la osteoporosis reducen significativamente el riesgo de fractura en mujeres con una densidad ósea muy baja (puntuaciones T menores de -2,5), con un efecto escaso o nulo en mujeres con mayor densidad ósea (p = 0,02 para la interacción) (9 , 10). En este caso, la inclusión únicamente de mujeres con una densidad ósea muy baja en un estudio puede aumentar la magnitud del efecto y reducir el tamaño de la muestra para un ensayo de tratamientos similares.

La estratificación de los participantes por una característica, como un grupo racial, permite a los investigadores incluir un número deseado de participantes con una característica que puede influir en el efecto del tratamiento o en su posibilidad de generalización. La reunión en un estrato se puede cerrar cuando se ha alcanzado el objetivo de participantes con esta característica. Sin embargo, como la mayoría de los estudios no están diseñados con un tamaño de la muestra suficiente para estudiar la heterogeneidad de los efectos de la intervención en dichos subgrupos, esta estrategia puede tener poca utilidad práctica.

Los criterios de exclusión deben ser parcos, porque las exclusiones innecesarias pueden dificultar la reunión del número necesario de participantes, reducen la posibilidad de generalizar los resultados, y aumentan la complejidad y el coste de esa reunión. Hay cinco razones principales para excluir personas de un ensayo clínico (tabla 10-1).

Se debe excluir a los posibles participantes si el tratamiento o la intervención del grupo testigo no son seguros. El tratamiento activo puede no ser seguro en personas que son susceptibles a los efectos adversos conocidos del tratamiento activo. Por ejemplo, el infarto de miocardio es un infrecuente efecto adverso del tratamiento con sildenafilo, por lo que los ensayos de este fármaco para tratar el vasoespasmo doloroso en pacientes con enfermedad de Raynaud deberán excluir a los pacientes con CI (11). Por el contrario, ser asignado al grupo de tratamiento inactivo o al grupo del placebo puede ser inseguro para algunos participantes. Por ejemplo, en mujeres con fracturas vertebrales, se sabe que los bisfosfonatos reducen el riesgo de fracturas posteriores, lo que hace que sea inaceptable que participen en un ensayo controlado con placebo de un nuevo tratamiento para la osteoporosis, salvo que también se administren bisfosfonatos a todas las participantes del ensayo. Debe excluirse a las personas en las que no es probable que el tratamiento activo sea eficaz, así como a aquellas que no es probable que acepten la intervención o realicen un seguimiento completo. En ocasiones, problemas prácticos, como una alteración del estado mental que dificulta el seguimiento de las instrucciones, justifican la exclusión. Los investigadores deben sopesar cuidadosamente los posibles criterios de exclusión que se aplican a muchas personas (p. ej., diabetes o límites superiores de edad), ya que pueden tener un gran efecto sobre la viabilidad y los costes de la reunión de participantes, y la posibilidad de generalizar los resultados.

Diseñar un tamaño de la muestra adecuado y planificar la reunión de participantes en consecuencia

Los ensayos clínicos con un número excesivamente bajo de participantes para detectar efectos importantes son ineficientes y poco éticos, y pueden conducir a conclusiones engañosas (12). Por lo tanto, calcular el tamaño de la muestra es una de las primeras y más importantes partes de la planificación de un ensayo (cap. 6), y se debe tener en consideración el hecho de que las tasas de respuesta en los ensayos clínicos suelen ser inferiores a las estimadas debido al sesgo del voluntario sano. Además, reunir participantes para un ensayo muchas veces es más difícil que hacerlo para un estudio de observación, porque los pacientes tienen que estar dispuestos a ser asignados aleatoriamente, muchas veces a un placebo o un fármaco «experimental». Por estos motivos, el investigador debe planificar una muestra abundante a partir de una gran población accesible, y el tiempo y el dinero suficientes para reunir el tamaño de la muestra deseado cuando (como suele ocurrir) los obstáculos para hacerlo se vuelven mayores de lo esperado.

MOTIVO	EJEMPLO: UN ENSAYO DE RALOXIFENO (UN MODULADOR SELECTIVO DEL RECEPTOR ESTROGÉNICO) O UN PLACEBO PARA EVITAR LA CARDIOPATÍA
1. Un tratamiento en estudio podría ser nocivo	
• Riesgo inaceptable de daños si se asigna al grupo de tratamiento activo	Episodio previo de tromboembolia venosa (el raloxifeno aumenta el riesgo de tromboembolia venosa)
• Riesgo inaceptable de daños si se asigna al grupo testigo	Reciente cáncer de mama positivo para receptores estrogénicos (el tratamiento con un modulador selectivo del receptor estrogénico es eficaz, y es el tratamiento estándar)
2. Es probable que el tratamiento activo no sea eficaz	
• Con riesgo bajo de la respuesta	Mujer adolescente con riesgo muy bajo de cardiopatía isquémica
• Tiene un tipo de enfermedad que no es probable que responda al tratamiento	Paciente con cardiopatía valvular que no es probable que responda a los efectos antiaterogénicos propuestos del raloxifeno
• Está recibiendo un tratamiento que es posible que interfiera con la intervención	Tratamiento con estrógenos (que compiten con el raloxifeno)
3. No es probable que cumpla con la intervención	Escaso cumplimiento durante el período de preinclusión (cap. 11)
4. No es probable que complete el seguimiento	Piensa mudarse antes de que acabe el estudio y no estará disponible para las mediciones finales de las variables de respuesta Esperanza de vida corta por una enfermedad grave
5. Problemas prácticos con la participación en el protocolo	Alteración del estado mental que impide una respuesta exacta a las preguntas

Volver al principio

▪ MEDICIÓN DE LAS VARIABLES BASALES

Para facilitar el contacto con participantes que se han perdido durante el seguimiento, es importante registrar los nombres, números de teléfono, direcciones y direcciones de correo electrónico de dos o tres amigos o familiares que siempre sepan cómo llegar al participante. Cuando sea permisible, también es valioso un registro de números de la seguridad social u otros números de identificación. Pueden

usarse para determinar la situación vital del participante y para detectar respuestas determinadas a través de registros sanitarios (p. ej., sistemas de seguros sanitarios). No obstante, esta es «información sanitaria protegida» confidencial, y no debe acompañar a datos que se envíen a un centro coordinador o una institución patrocinadora.

Descripción de los participantes

Los investigadores deben reunir información sobre los factores de riesgo y los posibles factores de riesgo de la respuesta, y sobre las características de los participantes que puedan afectar a la eficacia de la intervención. Estas mediciones también proporcionan un medio para comprobar la posibilidad de comparar los grupos del estudio aleatorizados en situación inicial y ofrecen información para evaluar la posibilidad de generalizar los resultados. El objetivo será asegurarse de que las diferencias de estas características iniciales no superan lo que podría esperarse por la acción del azar, lo que indicaría un error técnico o sesgo en la realización de la aleatorización. En estudios pequeños que son propensos a una llamativa desigualdad de la distribución de las características iniciales en los grupos aleatorizados únicamente por efecto del azar, la medición de factores predictivos importantes de la variable de respuesta permite el ajuste estadístico de la comparación aleatorizada para reducir la influencia de estas distribuciones desiguales, pero aleatorias. La medición de las variables predictivas de la variable de respuesta también permite que el investigador analice si la intervención tiene diferentes efectos en subgrupos clasificados por las variables iniciales (modificación del efecto; cap. 11).

Medición del valor inicial de la variable de respuesta

Si las respuestas incluyen el cambio de una variable, la variable de respuesta debe medirse al principio del estudio del mismo modo que se medirá al final. En los estudios con una respuesta dicotómica (los efectos de la terapia cognitiva conductual sobre las puntuaciones de depresión), la mejor medida generalmente es un cambio de la variable de respuesta durante el transcurso del estudio. Este método suele reducir al mínimo la variabilidad de la respuesta entre los participantes del estudio y ofrece mayor potencia que la simple comparación de los valores al final del estudio. En estudios que tienen una variable de respuesta dicotómica (p. ej., incidencia de CI) puede ser importante demostrar por la historia clínica y el electrocardiograma que la enfermedad no está presente al comienzo. También puede ser útil medir al comienzo las variables de respuestas secundarias y los resultados de los estudios auxiliares previstos.

Moderación

Tras haber señalado los múltiples usos de las determinaciones basales, debemos subrayar que el diseño de un ensayo clínico no necesita que se mida *todo*, porque la aleatorización minimiza el problema de la confusión por factores que ya existen al principio. Hacer muchas mediciones añade gastos y complejidad. En un estudio aleatorizado que tiene un presupuesto limitado, suele ser mejor gastar el tiempo y el dinero en cosas esenciales para la integridad del estudio, como la idoneidad del tamaño de la muestra, el éxito de la aleatorización y el enmascaramiento, la completitud del cumplimiento y el seguimiento. Yusuf y cols. han propuesto el uso de grandes ensayos con muy pocas mediciones (13).

Bancos de muestras

Guardar imágenes, suero, ADN, etc., al principio permitirá realizar mediciones posteriores de los cambios producidos por el tratamiento, marcadores biológicos que predicen la respuesta y factores como el genotipo que podrían identificar a personas que responden bien o mal al tratamiento. Las muestras almacenadas pueden ser también una importante fuente para estudiar otras preguntas de investigación no relacionadas directamente con la respuesta principal.

Volver al principio

▪ ALEATORIZACIÓN Y ENMASCARAMIENTO

El cuarto paso de la figura 10-1 es asignar de forma aleatoria los participantes a dos grupos. En el diseño más sencillo, un grupo recibe una intervención con un tratamiento activo y el otro recibe un placebo. La asignación aleatoria dispone que la edad, el sexo y otras características pronósticas del principio que podrían confundir una asociación observada (incluso las que se desconocen o no se han medido) se distribuirán homogéneamente, salvo por la variación aleatoria, entre los grupos aleatorizados en situación inicial. El enmascaramiento es importante para mantener la posibilidad de comparar los grupos del estudio durante el ensayo y para garantizar una evaluación no sesgada de los resultados.

Aleatorización

Debido a que la aleatorización es la pieza clave de un ensayo clínico, es importante que se realice correctamente. Las dos características más importantes son: que el procedimiento realmente distribuya los tratamientos al azar y que las asignaciones sean resistentes a la manipulación, de modo que ni factores intencionados ni factores no intencionados puedan influir en la aleatorización.

Es importante que al participante se le realicen las exploraciones iniciales, que se encuentre que es elegible para su inclusión y que consienta en participar en el estudio antes de que se realice la aleatorización. A continuación se le distribuye de forma aleatoria, mediante un algoritmo informático o mediante la aplicación de una serie de números aleatorios. Una vez que se genera una lista del orden aleatorio de asignación a los grupos del estudio, debe aplicarse a los participantes según una estricta secuencia de entrada en el ensayo.

Es esencial diseñar el procedimiento de asignación aleatoria de modo que los miembros del equipo de investigación no puedan influir en la distribución. Por ejemplo, en los estudios realizados en un centro, las asignaciones aleatorias del tratamiento pueden colocarse por adelantado en una serie de sobres sellados por alguien que no vaya a intervenir en su apertura. Todos los sobres deben tener un número (de modo que todos puedan contarse al final del estudio), ser opacos (para evitar que se transparenten) y ser resistentes a la manipulación. Cuando se distribuye al azar un participante, su nombre y el número del siguiente sobre sin abrir se registran primero *en presencia de un segundo miembro del equipo*, y ambas personas del equipo firman el sobre; *a continuación*, se abre el sobre y se apunta en un registro el grupo de tratamiento contenido en él y asignado al participante.

Los ensayos multicéntricos usan habitualmente un servicio de aleatorización aparte, a prueba de manipulación, con el que el personal del ensayo contacta cuando un participante elegible está listo para ser aleatorizado. El miembro del equipo proporciona el nombre y la ID del estudio al nuevo participante. Se registra esta información, y entonces se asigna de forma aleatoria el grupo de tratamiento utilizando un programa informático que proporciona un número de asignación al tratamiento ligado a las intervenciones. El tratamiento también puede asignarse aleatoriamente mediante programas informáticos en un solo centro de investigación, siempre y cuando estos programas sean resistentes a la manipulación. Se necesitan estrictas precauciones para impedir la manipulación de la aleatorización, porque los investigadores a veces se encuentran bajo presión para influir en ese proceso (p. ej., por una persona que parece particularmente adecuada para un grupo de tratamiento activo en un ensayo clínico controlado con placebo).

Técnicas especiales de aleatorización

El método preferido es típicamente la aleatorización simple de los participantes individuales a cada uno de los grupos de intervención. Los ensayos de tamaño pequeño o moderado tendrán un pequeño aumento de la potencia si se usan procedimientos de aleatorización especiales para equilibrar el número de participantes en cada uno de los grupos (aleatorización por bloques) y la distribución de variables iniciales que se sabe que predicen la respuesta (aleatorización estratificada por bloques).

La aleatorización por bloques es una técnica que se usa habitualmente para asegurar que el número de participantes se distribuya por igual entre los grupos del estudio. La aleatorización se realiza en «bloques» de tamaño predeterminado. Por ejemplo, si el tamaño del bloque es seis, la aleatorización se realiza normalmente dentro de cada bloque de seis hasta que se asigna aleatoriamente a tres personas a uno de los grupos, tras lo cual se asigna automáticamente a los participantes al otro grupo hasta que se completa el bloque de seis. Esto significa que, en un estudio de 30 participantes, se asignarán exactamente 15 a cada grupo, y en un estudio de 33 participantes, la desproporción no podría ser mayor de 18:15. La aleatorización por bloques con un tamaño de bloque fijo es menos adecuada para estudios sin enmascaramiento, porque la asignación al tratamiento de los participantes al final de cada bloque podría predecirse y manipularse. Este problema puede reducirse al mínimo variando aleatoriamente el tamaño de los bloques (p. ej., oscilando desde bloques de cuatro a ocho), según un programa que el investigador desconozca.

La aleatorización estratificada por bloques asegura que un importante factor pronóstico de la respuesta se distribuya más equitativamente entre los grupos del estudio de lo que dictaría el azar solo. En un ensayo del efecto de un fármaco para evitar fracturas, tener una factura vertebral previa es un factor pronóstico tan importante de la variable de respuesta que puede ser mejor asegurar que se asignan a cada grupo números similares de personas con fracturas vertebrales. Esto puede lograrse realizando la aleatorización por bloques por separado por «estratos» (los que tienen fracturas vertebrales y los que no). La aleatorización estratificada por bloques puede aumentar ligeramente la potencia de un ensayo pequeño al reducir la variación de la respuesta debido a desproporciones casuales en importantes variables basales. Tiene poca utilidad en ensayos de gran tamaño (más de 1 000 participantes), porque la asignación aleatoria asegura la distribución casi uniforme de las variables basales.

Una limitación importante de la aleatorización estratificada por bloques es el pequeño número de variables basales, no más de dos o tres, que pueden equilibrarse mediante esta técnica. Una técnica para abordar esta limitación es la aleatorización adaptativa, que utiliza una «moneda sesgada» para alterar la probabilidad de asignar a cada nuevo participante, de manera que, por ejemplo, una persona con una puntuación de riesgo elevado en cualquier número de variables pronósticas iniciales tendría una probabilidad ligeramente mayor de ser incluida en el grupo del estudio que tiene menor riesgo general basado en todos los participantes asignados aleatoriamente a ese punto. Entre las desventajas de esta técnica están la dificultad de explicar la probabilidad de asignación a los grupos del estudio a los posibles participantes durante el consentimiento informado y la complejidad de su aplicación, con un sistema computarizado interactivo que recalcula las probabilidades de la moneda sesgada con cada aleatorización.

Habitualmente, la mejor decisión es asignar números iguales a cada uno de los grupos en estudio, porque esto maximiza la potencia para cualquier tamaño de la muestra total dado. Sin embargo, la atenuación de la potencia con una proporción de incluso 2:1 es bastante

pequeña (14), y en ocasiones puede ser adecuada la asignación desigual de los participantes a los grupos de tratamiento y testigo (15):

- El aumento del cociente de participantes asignados al tratamiento activo respecto a los testigos puede hacer que el estudio sea más atractivo para los posibles participantes, como los que tienen infección por el VIH, que quisieran tener la mayor probabilidad de recibir un tratamiento activo si participan.

- La reducción del cociente de participantes que reciben el tratamiento activo respecto a los testigos puede hacer que el estudio sea asequible cuando la intervención es muy costosa (como en el estudio de dieta pobre en grasa *Women's Health Initiative* [16]).

- El aumento de la proporción asignada al grupo que actúa como testigo de varios grupos de tratamiento activo aumenta la potencia de cada una de las comparaciones, al aumentar la precisión de la estimación en el grupo testigo (como en el estudio *Coronary Drug Project* [17]).

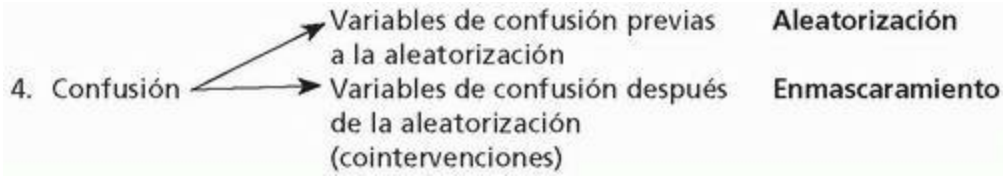
La aleatorización por parejas emparejadas es una estrategia para equilibrar variables de confusión basales que precisa la selección de parejas de participantes que coinciden en características importantes, como la edad y el sexo, y la asignación aleatoria posterior de un miembro de cada pareja a cada uno de los grupos del estudio. Un inconveniente de la aleatorización por parejas es que complica la reunión y la aleatorización de participantes, pues precisa que se demore la aleatorización de un participante elegible hasta que se identifique una pareja adecuada. Además, el emparejamiento no suele ser necesario en ensayos grandes en los que la asignación aleatoria equilibra los grupos según las variables iniciales. No obstante, puede usarse una versión atractiva de este diseño cuando las circunstancias permiten un contraste de los efectos de tratamiento y control en dos partes del mismo participante. En el *Diabetic Retinopathy Study*, por ejemplo, cada participante tenía un ojo asignado aleatoriamente a tratamiento de fotocoagulación, y el otro ojo actuaba como testigo (18).

Enmascaramiento

Siempre que sea posible, el investigador debe diseñar la intervención de tal forma que los participantes en el estudio, el personal que está en contacto con ellos, las personas que realizan las determinaciones y las que evalúan y adjudican las respuestas desconozcan la asignación del grupo del estudio. Cuando no es posible ocultar la asignación a todas estas personas, es muy deseable hacerlo con el mayor número posible (p. ej., ocultar siempre la asignación al personal que realice las mediciones de la variable de respuesta). En un ensayo aleatorizado, el enmascaramiento*es tan importante como* la aleatorización. La aleatorización minimiza la influencia de variables de confusión en el momento de la aleatorización, pero no influye sobre las diferencias que surgen entre los grupos durante el período de seguimiento (tabla 10-2). El enmascaramiento minimiza las fuentes de sesgo posteriores a la aleatorización, como las cointervenciones y el sesgo en la evaluación y la adjudicación de las variables de respuesta.

Ya se ha discutido el uso del enmascaramiento para evitar el sesgo producido por las cointervenciones (fármacos, otros tratamientos o conductas distintos a la intervención del estudio que modifican el riesgo de presentar la respuesta de interés) (pág. 139). El segundo objetivo importante del enmascaramiento es minimizar el sesgo en la evaluación y la adjudicación de la respuesta. En un estudio sin enmascaramiento, el investigador puede estar tentado de buscar con mayor cuidado las respuestas en el grupo no tratado o en diagnosticar la respuesta con mayor frecuencia. Por ejemplo, en un ensayo sin enmascaramiento del tratamiento con estatinas, puede ser más probable que los investigadores pregunten a los participantes del grupo de tratamiento activo sobre el dolor muscular espontáneo o a la presión, y que soliciten pruebas para hacer el diagnóstico de miositis. El enmascaramiento de los pacientes es particularmente importante cuando las variables de respuesta se basan en síntomas referidos por el propio participante.

EXPLICACIÓN DE LA ASOCIACIÓN	ESTRATEGIA PARA DESCARTAR LA EXPLICACIÓN RIVAL
1. Azar	Igual que en estudios de observación (v. tabla 9-2)
2. Sesgo	Igual que en estudios de observación (v. tabla 9-2)
3. Efecto-causa	(No es una explicación posible en un ensayo)



5. Causa-efecto

Una vez que se ha comprobado un posible episodio de respuesta, puede hacer falta su adjudicación. Por ejemplo, si la variable de respuesta del estudio es el infarto de miocardio, los investigadores habitualmente recogen datos sobre los síntomas, los hallazgos del ECG y las enzimas cardíacas. Posteriormente, expertos que desconocen el grupo de tratamiento utilizan estos datos y definiciones específicas para adjudicar si se ha producido o no un infarto de miocardio. Los resultados del ensayo C*anadian Cooperative Multiple Sclerosis* ilustran la importancia del enmascaramiento en la adjudicación no sesgada de respuesta (19). Se asignó aleatoriamente a pacientes con esclerosis múltiple a recibir: plasmaféresis combinada, ciclofosfamida y prednisona, o plasmaféresis simulada y placebos de los medicamentos. Al final del ensayo se evaluó la gravedad de la esclerosis múltiple mediante una exploración estructurada realizada por neurólogos que desconocían la asignación del tratamiento, y de nuevo por neurólogos que sí conocían dicha asignación. El tratamiento no fue eficaz según la evaluación de los neurólogos que desconocían la asignación, pero tuvo una eficacia estadísticamente significativa de acuerdo con la evaluación de los neurólogos que sí conocían la asignación. Los neurólogos que conocían la asignación no estaban intentando sesgar a propósito el resultado del estudio, sino que hay un intenso deseo humano de ver cómo mejoran los pacientes después del tratamiento, especialmente si el tratamiento es doloroso o potencialmente perjudicial. El enmascaramiento minimiza esta adjudicación sesgada de las variables de respuesta.

La evaluación con enmascaramiento de la respuesta puede ser menos importante si la respuesta del ensayo es una respuesta objetiva sólida, como la muerte o una medición automática sobre la que hay pocas o ninguna duda ni oportunidad de una evaluación sesgada. La mayor parte de las demás respuestas, como muerte por una causa específica, diagnóstico de una enfermedad, determinaciones físicas, escalas de cuestionarios y afecciones informadas por uno mismo, pueden verse afectadas por una comprobación y una adjudicación sesgadas.

Una vez que el ensayo ha finalizado, puede ser una buena idea evaluar si los participantes y los investigadores realmente desconocían la asignación, pidiéndoles que adivinaran a qué tratamiento se asignó al participante. Si una proporción mayor de la esperada adivina correctamente, la conclusión publicada de los hallazgos debe incluir la evaluación de los posibles sesgos que puede haber causado la ausencia parcial de enmascaramiento.

¿Qué hacer cuando el enmascaramiento es imposible?

En algunos casos, el enmascaramiento es difícil o imposible, bien sea por razones técnicas o éticas. Por ejemplo, es difícil conseguir que los participantes desconozcan la asignación si se les asigna a una intervención educativa, dietética o de ejercicio. Con frecuencia, las intervenciones quirúrgicas no pueden enmascararse, porque puede no ser ético realizar una falsa intervención en el grupo testigo. Sin embargo, la cirugía lleva siempre algún riesgo asociado, por lo que es muy importante determinar si el procedimiento es realmente eficaz. Por ejemplo, en un reciente estudio aleatorizado se observó que el desbridamiento artroscópico del cartílago de la rodilla no era más eficaz que la artroscopia con desbridamiento simulado para aliviar el dolor artrósico de la rodilla (20). En este caso, el riesgo de los pacientes del grupo testigo podría haberse compensado si los resultados del estudio hubieran evitado que a miles de pacientes se les hubiera realizado un procedimiento ineficaz.

Si las intervenciones no pueden enmascararse, el investigador debe, al menos, limitar lo más posible las cointervenciones, y garantizar que las personas que evalúen y adjudiquen las variables de respuesta desconozcan la asignación. Por ejemplo, un investigador que investigara el yoga para el alivio de los sofocos menopáusicos podría indicar a las participantes en el grupo de yoga y del grupo testigo que no iniciaran nuevos fármacos, actividades de relajación ni ningún otro tratamiento para los sofocos hasta que el estudio hubiera finalizado. También, el personal del estudio que recopilara información sobre la intensidad de los sofocos podría ser diferente del que imparte las clases de yoga.

Volver al principio

▪ RESUMEN

- Un estudio aleatorizado con enmascaramiento, diseñado y realizado adecuadamente, puede ofrecer la inferencia causal más

definitiva como base para las directrices de práctica clínica sustentadas por la medicina basada en las pruebas.

- La elección y la dosis de la intervención es una decisión difícil que equilibra juicios sobre la eficacia y la seguridad; otras consideraciones son la relevancia para la práctica clínica, la idoneidad para el enmascaramiento y si se debe usar una combinación de fármacos.

- Cuando sea posible, el grupo de comparación debe ser un grupo testigo tratado con placebo que permita el enmascaramiento de los participantes, los investigadores y el personal del estudio.

- Las respuestas clínicamente importantes, como el dolor, la calidad de vida, la incidencia de cáncer y la muerte, son las respuestas más significativas de los ensayos. Las variables de respuesta intermedias, como la viremia del VIH, son marcadores indirectos válidos para respuestas clínicas en la medida en que los cambios del marcador inducidos por el tratamiento predicen los cambios de la respuesta clínica.

- La medición de más de una variable de respuesta habitualmente es útil, aunque su combinación en variables de respuesta compuestas precisa una consideración cuidadosa; se debe especificar una variable de respuesta principal única para estudiar la hipótesis principal.

- Todos los ensayos clínicos contienen medidas de posibles efectos adversos de la intervención, tanto dirigidas como (con moderación) abiertas, con procedimientos para garantizar que se notifiquen rápidamente los acontecimientos adversos graves (AAG) a los CRI y a los promotores.

- Los criterios para seleccionar a los participantes en el estudio deben identificar a los que tienen probabilidad de experimentar el máximo efecto beneficioso y el mínimo efecto perjudicial por el tratamiento, y de cumplir los protocolos de tratamiento y seguimiento. La elección de participantes con riesgo elevado de la variable de respuesta puede reducir el tamaño de la muestra, aunque puede dificultar la inclusión de pacientes y reducir la posibilidad de generalizar los hallazgos.

- Las variables basales deben medirse de forma parca para describir las características de los participantes, medir factores de riesgo y valores basales de la variable de respuesta, y facilitar el examen posterior de los diferentes efectos de la intervención en varios subgrupos (modificación del efecto). Se debe plantear el almacenamiento de suero, material genético, imágenes, etc., en situación inicial, para su análisis posterior.

- La aleatorización, que minimiza la influencia de las variables de confusión en situación inicial, debe protegerse de la alteración; la aleatorización por parejas es un diseño excelente cuando es posible, y en los estudios pequeños la aleatorización por bloques estratificada puede disminuir la posibilidad de distribuciones desiguales de variables predictivas fundamentales.

- El enmascaramiento de la intervención es tan importante como la aleatorización, y sirve para controlar las cointervenciones y los sesgos en la evaluación, y la adjudicación de las variables de respuesta.

Volver al principio

BIBLIOGRAFÍA

1. Ettinger B, Black DM, Mitlak BH, et al. Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene: results from a 3-year randomized clinical trial. Multiple Outcomes of Raloxifene Evaluation (MORE) investigators. JAMA 1999;282:637-645. Citado aquí

2. The Women's Health Initiative Study Group. Design of the women's health initiative clinical trial and observational study. Control Clin Trials 1998;19:61-109. Citado aquí

3. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 1989;8:431-440. Citado aquí

4. Cummings SR, Karpf DB, Harris F, et al. Improvement in spine bone density and reduction in risk of vertebral fractures during treatment with antiresorptive drugs. Am J Med 2002;112:281-289. Citado aquí

5. Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. JAMA 1998;280:605-613. Citado aquí

6. Cummings SR, San Martin J, McClung MR, et al. Denosumab for prevention of fractures in postmenopausal women with

osteoporosis. N Engl J Med 2009;361(8):756-765. Citado aquí

7. Mosca L, Barrett-Connor E, Wenger NK, et al. Design and methods of the Raloxifene Use for The Heart (RUTH) Study. Am J Cardiol 2001;88:392-395. Citado aquí

8. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the women's health initiative randomized controlled trial. JAMA 2002; 288:321-333. Citado aquí

9. McClung M, Boonen S, Torring O, et al. Effect of denosumab treatment on the risk of fractures in subgroups of women with postmenopausal osteoporosis. J Bone Miner Res 2011;27:211-218. Citado aquí

10. Cummings SR, Black DM, Thompson DE, et al. Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the fracture intervention trial. JAMA 1998;280:2077-2082. Citado aquí

11. Fries R, Shariat K, von Wilmowsky H, et al. Sildenafil in the treatment of Raynaud's phenomenon resistant to vasodilatory therapy. Circulation 2005;112:2980-2985. Citado aquí

12. Freiman JA, Chalmers TC, Smith H Jr, et al. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 “negative” trials. N Engl J Med 1978;299:690-694. Citado aquí

13. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? Stat Med 1984;3:409-420. Citado aquí

14. Friedman LM, Furberg C, DeMets DL. Fundamentals of clinical trials, 4th ed. New York: Springer, 2010. Citado aquí

15. Avins AL. Can unequal be more fair? Ethics, subject allocation, and randomised clinical trials. J Med Ethics 1998;24:401-408. Citado aquí

16. Prentice RL, Caan B, Chlebowski RT, et al. Low-fat dietary pattern and risk of invasive breast cancer: the women's health initiative randomized controlled dietary modification trial. JAMA 2006;295:629-642. Citado aquí

17. CDP Research Group. The coronary drug project. Initial findings leading to modifications of its research protocol. JAMA 1970;214:1303-1313. Citado aquí

18. Diabetic Retinopathy Study Research Group. Preliminary report on effects of photocoagulation therapy. Am J Ophthalmol 1976;81:383-396. Citado aquí

19. Noseworthy JH, O'Brien P, Erickson BJ, et al. The Mayo-Clinic Canadian cooperative trial of sulfasalazine in active multiple sclerosis. Neurology 1998;51:1342-1352. Citado aquí

20. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. N Engl J Med 2002;347:81-88. Citado aquí

CAPÍTULO 11 Otros diseños de ensayos clínicos y problemas en su realización

Deborah Grady
Steven R. Cummings
Stephen B. Hulley

En el último capítulo se comentó el clásico ensayo clínico aleatorizado, con enmascaramiento y de grupos paralelos: cómo elegir y enmascarar las situaciones de intervención y testigo, asignar aleatoriamente las intervenciones, escoger las respuestas, manejar los acontecimientos adversos, seleccionar los participantes y medir las variables basales y las variables de respuesta.

En este capítulo se describen otros diseños de estudios comparativos entre grupos, tanto aleatorizados como no aleatorizados, además de diseños intragrupo, estudios cruzados y estudios piloto. Después se aborda la realización de los estudios clínicos, con aspectos como el cumplimiento de la intervención y el seguimiento, y la evaluación y la adjudicación de las respuestas. Se concluye con un análisis de problemas estadísticos, como el seguimiento provisional para la finalización temprana del estudio, el análisis de la intención de tratar y según protocolo, y el uso de análisis de subgrupos para descubrir la modificación del efecto.

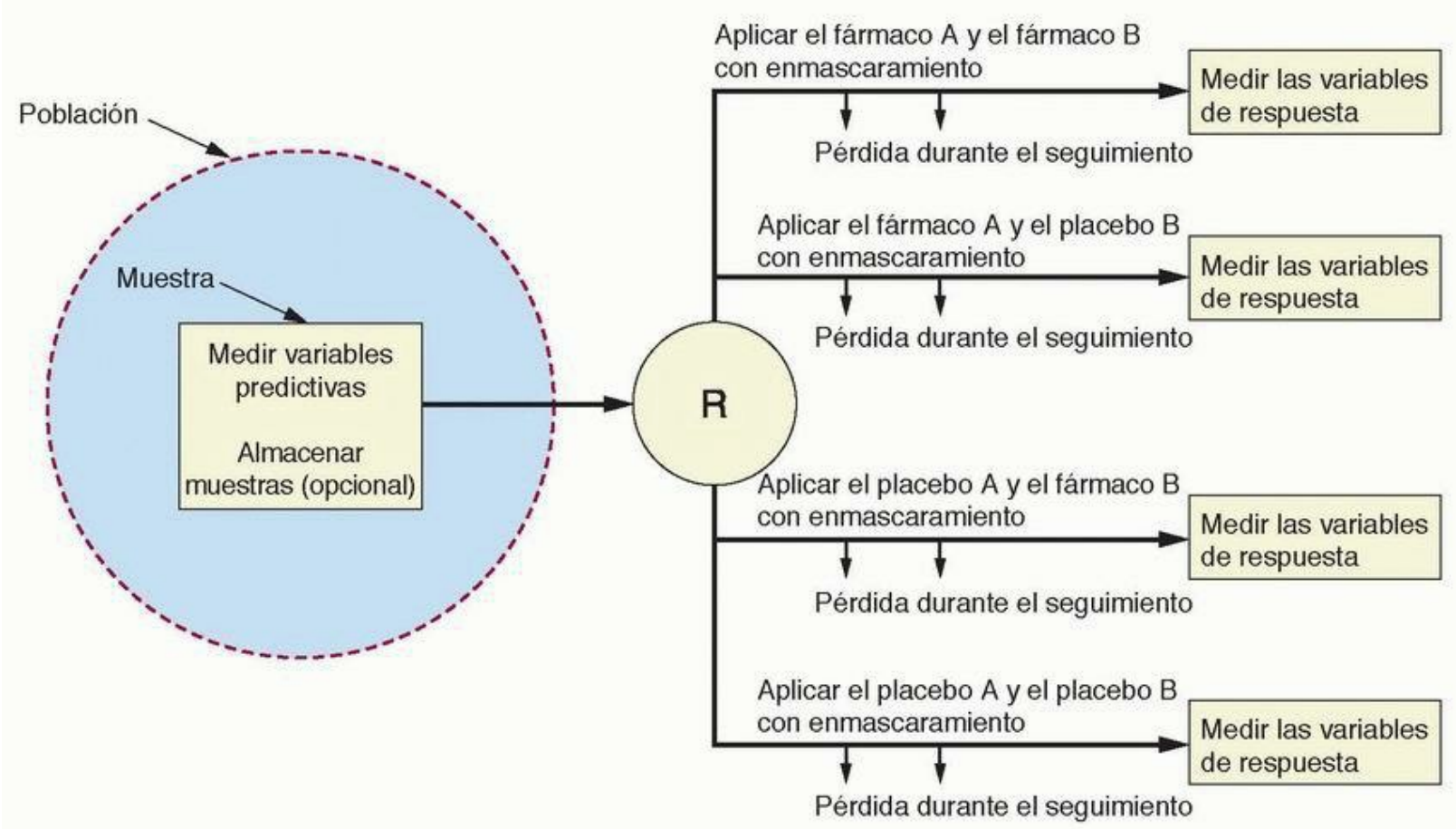
▪ OTROS DISEÑOS ALEATORIZADOS

Existen diversas variaciones del clásico ensayo clínico aleatorizado de grupos paralelos que pueden ser útiles cuando las circunstancias son las adecuadas.

Diseño factorial

El diseño factorial pretende responder a dos (o más) preguntas de investigación separadas en un solo estudio (fig. 11-1). Un buen ejemplo es el *Women's Health Study*, que se diseñó para evaluar el efecto del ácido acetilsalicílico en dosis bajas y la vitamina E sobre el riesgo de sufrir episodios cardiovasculares en mujeres sanas (1). Se distribuyó aleatoriamente a las participantes en cuatro grupos, y se analizaron dos hipótesis comparando dos mitades de la cohorte en estudio. En primer lugar se comparó la incidencia de episodios cardiovasculares en mujeres tratadas con ácido acetilsalicílico con la de mujeres tratadas con un placebo del ácido acetilsalicílico (sin tener en cuenta el hecho de que la mitad de cada uno de estos grupos recibió vitamina E); a continuación se comparó la incidencia de episodios cardiovasculares en las mujeres tratadas con vitamina E con la de las que recibieron un placebo de vitamina E (sin tener en cuenta ahora que la mitad de cada uno de estos grupos recibió ácido acetilsalicílico). Los investigadores tienen dos estudios completos por el precio de uno.

Una limitación es la posibilidad de modificación del efecto (interacción): si el efecto del ácido acetilsalicílico sobre el riesgo de enfermedad cardiovascular es diferente en las mujeres tratadas con vitamina E y en las mujeres no tratadas con vitamina E, entonces hay modificación del efecto, y se debería calcular por separado el efecto del ácido acetilsalicílico en estos dos grupos. Esto reduciría la potencia de estas comparaciones, porque solo se incluiría en cada análisis a la mitad de las participantes. Los diseños factoriales pueden usarse, realmente, para *estudiar* la modificación del efecto, pero los estudios diseñados para ello son más complicados y difíciles de realizar, se precisan tamaños de la muestra mayores y los resultados pueden ser difíciles de interpretar. Otras limitaciones del diseño factorial son que la misma población de estudio debe ser adecuada para cada una de las intervenciones, y que múltiples tratamientos pueden interferir con la inclusión de los participantes y su cumplimiento, y los análisis son más complejos. Una vez dicho esto, el diseño factorial puede ser muy eficiente. Por ejemplo, en el ensayo aleatorizado *Women's Health Initiative* se pudo comprobar el efecto de tres intervenciones (hormonoterapia posmenopáusica, dieta baja en grasas y calcio más vitamina D) sobre varias respuestas (2).



▪FIGURA 11-1. En un estudio aleatorizado factorial, los pasos son:
• Seleccionar una muestra de participantes de una población idónea para recibir la intervención.
• Medir las variables predictivas y (si procede) el nivel inicial de la variable de respuesta.
• Considerar la opción de almacenar suero, imágenes, etc., para su análisis posterior.
• Asignar aleatoriamente dos (o más) intervenciones activas y sus testigos a cuatro (o más) grupos.
• Seguir las cohortes a lo largo del tiempo, minimizando la pérdida durante el seguimiento y evaluando el cumplimiento de las condiciones de intervención y testigo.
• Medir las variables de respuesta.
• Analizar los resultados, comparando primero los dos grupos de la intervención A (combinados) con los grupos del placebo A combinados, y después comparando los dos grupos de la intervención B (combinados) con los grupos del placebo B combinados.
Herramientas de imágenes

Aleatorización por conglomerados

La aleatorización por conglomerados necesita que el investigador asigne de forma aleatoria a las intervenciones a grupos o conglomerados de participantes que se producen de forma natural, en lugar de asignarlas a individuos. Un buen ejemplo sería un ensayo en el que se incluyó a jugadores de 120 equipos universitarios de béisbol, se distribuyó al azar a la mitad de los equipos a una intervención en la que se fomentaba el abandono del uso del tabaco en polvo, y se observó una frecuencia significativamente menor de este uso en los jugadores de los equipos que recibieron la intervención que en los equipos testigos (3). Aplicar la intervención a grupos de personas puede ser más viable y rentable que tratar a los individuos de uno en uno, y se pueden abordar mejor preguntas de investigación sobre los efectos de programas de salud pública en la población. Algunas intervenciones, como una dieta con pocas grasas, son difíciles de realizar en un solo miembro de una familia. Cuando se aleatoriza individualmente a los participantes de un grupo natural, es probable que los que reciben la intervención comenten o compartan esta con los miembros de la familia, compañeros, miembros del equipo o conocidos que han sido asignados al grupo testigo.

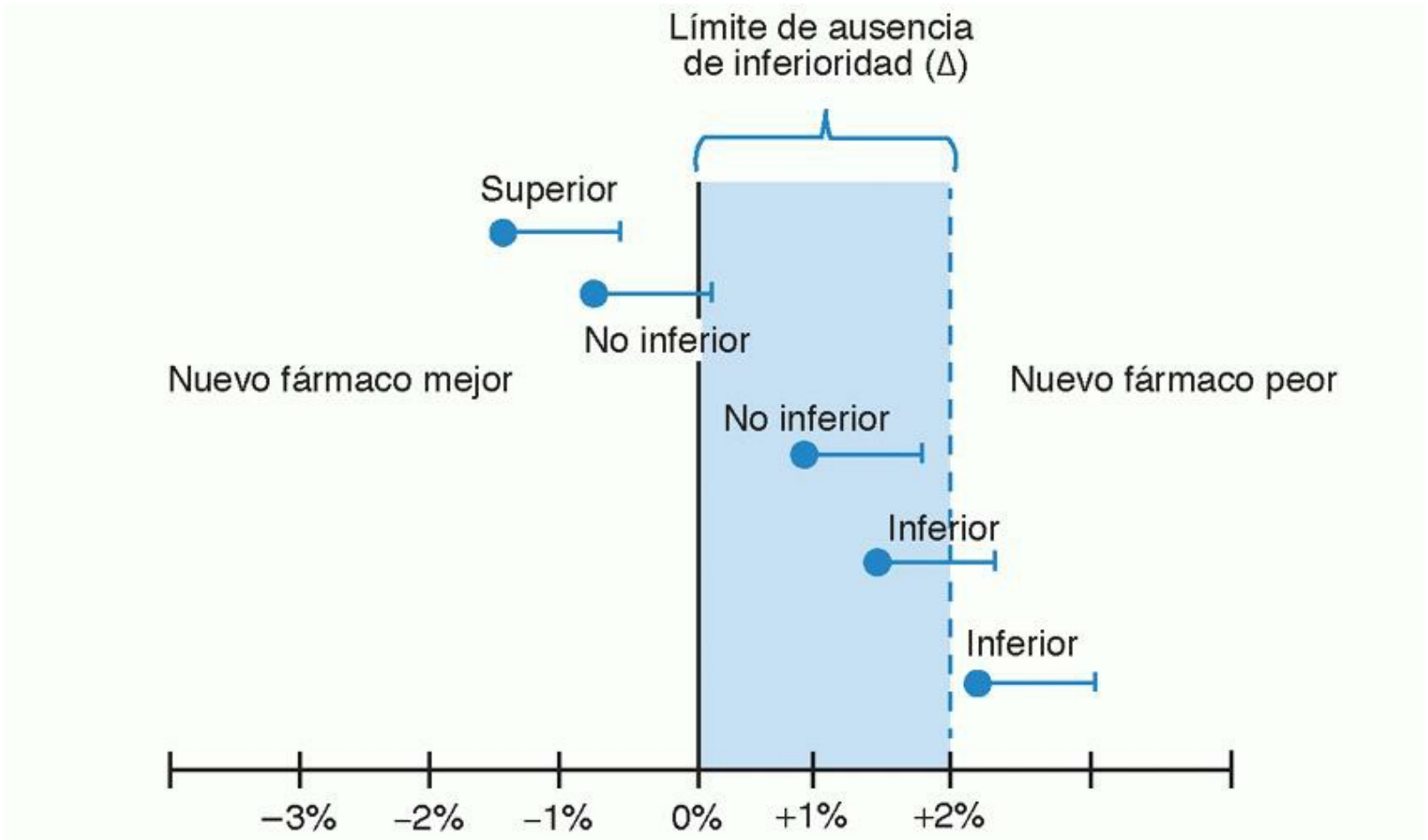
En el diseño de aleatorización por conglomerados, las unidades de aleatorización y análisis son grupos, no individuos. Por lo tanto, el tamaño de la muestra eficaz es menor que el número de participantes individuales y la potencia es menor. El tamaño eficaz de la muestra depende de la correlación del efecto de la intervención en los participantes de los conglomerados, y se encuentra en algún punto entre el número de grupos y el número de participantes (4). Otro inconveniente es que el cálculo del tamaño de la muestra y el análisis de los datos son más complicados en los diseños aleatorizados por conglomerados que en los de aleatorización individual (4).

Estudios con testigo activo: equivalencia y ausencia de inferioridad

Un estudio con testigo activo es un estudio en el que el grupo testigo recibe un tratamiento activo. Este diseño puede ser óptimo cuando hay un tratamiento eficaz conocido o «tratamiento estándar» para una enfermedad. En ocasiones, a este tipo de estudio se le denomina estudio de eficacia comparativo, porque se comparan dos tratamientos.

En algunos casos, el objetivo de un estudio con testigo activo es mostrar que un nuevo tratamiento es superior a un tratamiento de eficacia establecida. En esa situación, el diseño y los métodos son similares a los de un estudio controlado con placebo. Sin embargo, en la mayoría de los casos, los investigadores quieren establecer que un nuevo tratamiento que tiene algunas ventajas respecto a un tratamiento de eficacia establecida (más fácil de utilizar, menos invasivo, más seguro) tiene una eficacia *similar*. En este caso es más adecuado un estudio de equivalencia o de ausencia de inferioridad.

Los métodos estadísticos para los estudios de equivalencia y ausencia de inferioridad son diferentes de los métodos de los estudios diseñados para demostrar que un tratamiento es mejor que otro. En un estudio diseñado para mostrar que un tratamiento es superior, en el análisis habitual se utilizan pruebas de significación estadística para aceptar o rechazar la hipótesis nula de que no hay diferencias entre los grupos. En un estudio diseñado para mostrar que un nuevo tratamiento es equivalente al tratamiento estándar, por otro lado, el objetivo ideal sería *aceptar* la hipótesis nula de que no hay diferencias. Sin embargo, la demostración de que *no* hay diferencias entre los tratamientos (ni siquiera una diferencia minúscula) precisaría un tamaño de la muestra infinito. Por ello, la solución práctica es diseñar el tamaño de la muestra y el plan del análisis utilizando un abordaje de intervalos de confianza (IC), considerando el lugar en el que se encuentra el IC del efecto de un nuevo tratamiento en comparación con el tratamiento estándar en relación con un valor delta («Δ») preespecificado, la diferencia inaceptable de eficacia entre los dos tratamientos (5 , 6). Se considera que se ha establecido la equivalencia o la ausencia de inferioridad al nivel de significación especificado por el IC si el IC alrededor de la diferencia entre la eficacia del tratamiento nuevo y la del tratamiento establecido no incluye Δ (fig. 11-2). Se trata de un abordaje bilateral en el caso de un estudio de equivalencia (es decir, el nuevo tratamiento no es peor *ni* mejor que el tratamiento estándar). Sin embargo, es infrecuente que los investigadores estén interesados en si un nuevo tratamiento es *a la vez* no mejor *y* no peor que un tratamiento ya establecido. La mayoría de los veces los investigadores están especialmente interesados en mostrar que un nuevo tratamiento con otras ventajas no es inferior al tratamiento estándar. La naturaleza unilateral del diseño del estudio de ausencia de inferioridad también tiene la ventaja de permitir un tamaño de la muestra más pequeño o un menor valor de a habitualmente se prefiere este último (p. ej., 0,025 en lugar de 0,05), es decir, un valor más conservador.



Límites inferiores del intervalo de confianza del 95 % para las diferencias de tratamiento por incidencia de accidente cerebrovascular en pacientes con fibrilación auricular aleatorizados para el tratamiento con warfarina u otro fármaco

▪FIGURA 11-2. Posibles resultados en un estudio de ausencia de inferioridad en el que se comparó un nuevo fármaco con warfarina como tratamiento para reducir el riesgo de accidente cerebrovascular en pacientes con fibrilación auricular, definiendo el límite de ausencia de inferioridad (delta) en +2 %. Se muestran los intervalos de confianza del 95 % unilaterales alrededor de la diferencia de la incidencia de accidentes cerebrovasculares entre la warfarina y el nuevo fármaco, y se ilustran los resultados de superioridad, inferioridad y ausencia de inferioridad.
Herramientas de imágenes

Uno de los problemas más exigibles en el diseño de un estudio de ausencia de inferioridad es establecer el límite de ausencia de inferioridad (Δ), la pérdida de eficacia del nuevo tratamiento que sería inaceptable (7). Esa decisión se basa en consideraciones estadísticas y clínicas de la posible eficacia y las ventajas del nuevo tratamiento, y es necesario un juicio experto (8) (en el apéndice 11A se puede ver un ejemplo de cómo se realiza esto). En general, los estudios de ausencia de inferioridad deben ser más extensos que los estudios controlados con placebo, porque la diferencia aceptable entre el tratamiento nuevo y el tratamiento establecido habitualmente es menor que la diferencia esperada entre un nuevo tratamiento y el placebo.

Debe señalarse que la ausencia de inferioridad puede no significar que el tratamiento establecido y el tratamiento nuevo sean eficaces: podrían ser equivalentes en cuanto a la ineficacia o el carácter lesivo. Para asegurarse de que un nuevo tratamiento evaluado en un estudio de ausencia de inferioridad es más eficaz que el placebo, debe haber datos sólidos previos que respalden la eficacia del tratamiento establecido. Esto también significa que el diseño del estudio de ausencia de inferioridad debe ser lo más similar posible a los estudios que han determinado la eficacia del tratamiento estándar, con similitud de aspectos como criterios de selección, dosis del tratamiento establecido, cumplimiento del régimen estándar, duración del seguimiento, pérdidas durante el seguimiento, etc. (6 , 7). Cualquier problema que reduzca la eficacia del tratamiento estándar (inclusión de participantes que tengan poca probabilidad de beneficiarse, incumplimiento del tratamiento, pérdida durante el seguimiento) aumentará la probabilidad de que se encuentre que el nuevo tratamiento no es inferior, simplemente porque la eficacia del tratamiento estándar se ha reducido. Puede parecer que un tratamiento nuevo y menos eficaz no es inferior cuando, en realidad, los hallazgos representan un estudio mal realizado.

En resumen, los estudios de ausencia de inferioridad y de equivalencia son particularmente útiles si un nuevo tratamiento tiene ventajas importantes, como menor coste, facilidad de uso o mayor seguridad. Es difícil justificar estudios extensos para evaluar un nuevo fármaco de «yo también» sin ninguna de estas ventajas. Debe insistirse en que los estudios de ausencia de inferioridad y de equivalencia pueden llevar a la conclusión errónea de que los tratamientos son equivalentes si el estudio está mal realizado.

Diseños adaptativos

Los estudios clínicos generalmente se realizan cumpliendo un protocolo que no se modifica durante la realización del estudio. Sin embargo, para algunos tipos de tratamientos de enfermedades es posible hacer un seguimiento de los resultados del estudio a medida que avanza y modificar el diseño del estudio de acuerdo con el análisis provisional de los resultados (9). Por ejemplo, considere un estudio de varias dosis de un nuevo tratamiento para la dispepsia no ulcerosa. Según el diseño inicial, se puede planificar la inclusión de 50 participantes en un grupo del placebo y 50 en cada uno de tres grupos de dosis durante 12 semanas de tratamiento, con un período de inclusión de 1 año de duración. La revisión de los resultados después de que los primeros 10 participantes de cada grupo hayan finalizado 4 semanas de tratamiento podría mostrar que hay una tendencia hacia el alivio de la dispepsia solo en el grupo de mayor dosis. Puede ser más eficiente finalizar la asignación de participantes a las dos dosis menores, y continuar la aleatorización solo a los grupos de mayor dosis y del placebo. Otros aspectos de un estudio que se podrían modificar basándose en los resultados provisionales son el aumento o la disminución del tamaño de la muestra o de la duración del estudio si los resultados provisionales indican que la magnitud del efecto o la incidencia de las respuestas difieren de los supuestos originales.

Los diseños adaptativos son posibles únicamente con tratamientos que producen respuestas que se miden y analizan en fases suficientemente tempranas del transcurso del estudio para hacer que sean posibles los cambios de diseño en fases posteriores del estudio. Para evitar el sesgo, se deben establecer, antes del comienzo del estudio, reglas sobre cómo se puede modificar el diseño, y los análisis provisionales y el planteamiento del cambio del diseño los debe realizar un comité de supervisión de los datos y de la seguridad independiente que revise los datos sin enmascaramiento. Los análisis provisionales múltiples aumentarán la probabilidad de encontrar un resultado favorable que se deba a variación aleatoria y, cuando se analicen los resultados, se debe tener en consideración la mayor probabilidad de error de tipo I.

Además de ser más complejos de realizar y de analizar, los diseños adaptativos precisan que el consentimiento informado incluya todos los posibles cambios del diseño del estudio, y es difícil estimar el coste de un estudio adaptativo y de los recursos específicos necesarios para realizarlo. A pesar de esas precauciones y limitaciones, los diseños adaptativos son eficientes y pueden ser útiles, especialmente durante el desarrollo de un nuevo tratamiento; pueden permitir la identificación más temprana de las mejores dosis y duración del tratamiento, y garantizan que una proporción elevada de los participantes reciban el tratamiento óptimo.

Volver al principio

▪ DISEÑOS NO ALEATORIZADOS

Diseños no aleatorizados entre grupos

Los ensayos clínicos que comparan grupos que no han sido aleatorizados son mucho menos eficaces que los aleatorizados para controlar las variables de confusión. Por ejemplo, en un estudio en el que se comparan los efectos de la cirugía de derivación arterial coronaria con los de la angioplastia percutánea, si se permite que los médicos decidan a qué pacientes van a realizar los procedimientos, en lugar de utilizar asignación aleatoria, es probable que los pacientes asignados a cirugía sean distintos de los pacientes asignados a angioplastia. Los métodos analíticos pueden ajustarse para factores basales que son desiguales en los dos grupos del estudio, pero esta estrategia no aborda el problema de las variables de confusión no medidas. Cuando se comparan los hallazgos de estudios aleatorizados y no aleatorizados para responder a una misma pregunta de la investigación, los beneficios aparentes de la intervención son a menudo mayores en los estudios no aleatorizados, incluso después de ajustar estadísticamente las diferencias en las variables basales (10). El problema de la confusión en los estudios clínicos no aleatorizados puede ser grave, y puede que no se elimine completamente mediante el ajuste estadístico (11).

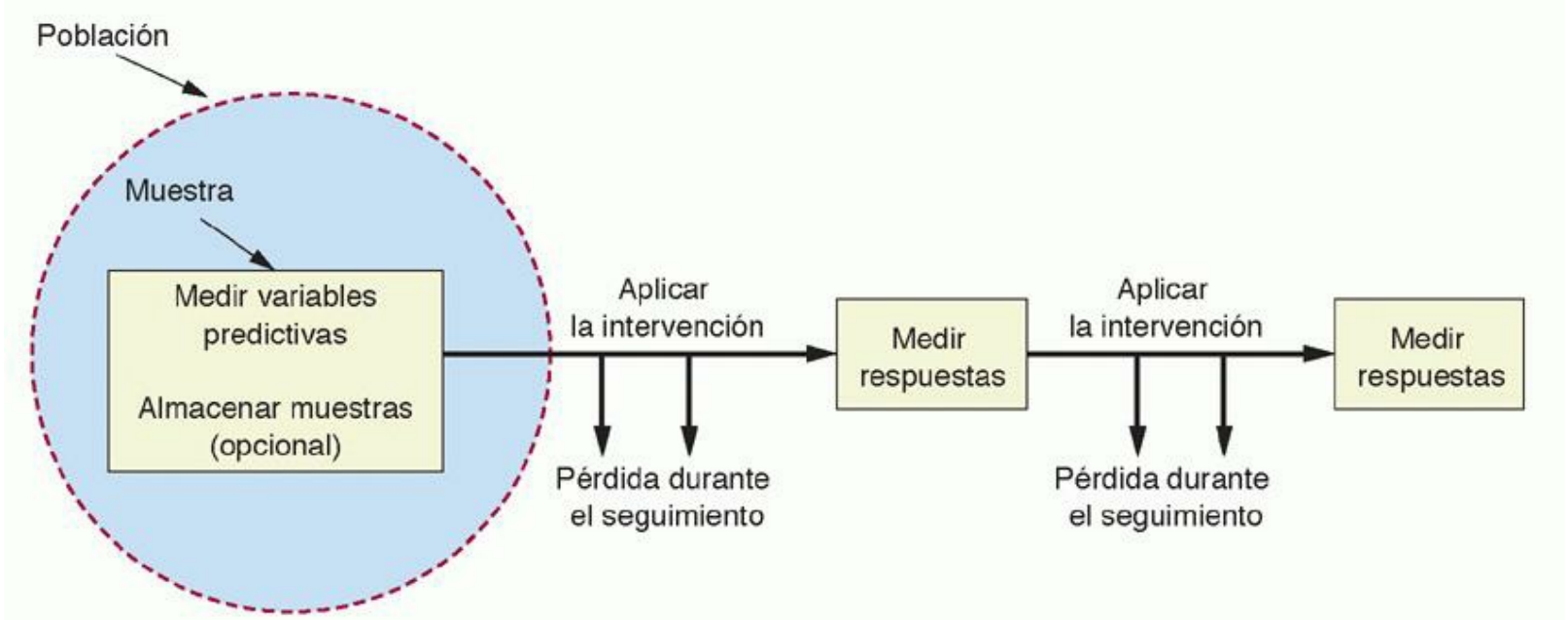
A veces se distribuyen los participantes a los grupos de estudio mediante un mecanismo seudoaleatorio. Por ejemplo, se puede asignar a los participantes con un número de historia al grupo de tratamiento. Estos diseños pueden ofrecer ventajas logísticas, pero la posibilidad de predecir la asignación al grupo de estudio permite al investigador o al personal del estudio alterarlo, manipulando la secuencia o la idoneidad de nuevos participantes.

A veces, el investigador asigna los participantes a los grupos de estudio según determinados criterios específicos. Por ejemplo, puede asignarse a los pacientes con diabetes a recibir insulina cuatro veces al día o a recibir insulina de acción prolongada una vez al día, según su disposición a aceptar cuatro inyecciones diarias. El problema de este diseño es que los que aceptan cuatro inyecciones al día podrían ser diferentes de aquellos que no quieren (p. ej., cumplir mejor los demás consejos sanitarios), y esta podría ser la causa de cualquier diferencia observada en las respuestas de los dos programas de tratamiento.

Los diseños no aleatorizados se escogen a veces con la creencia errónea de que son más éticos que la aleatorización, porque permiten que el participante o el médico elijan la intervención. De hecho, los estudios solo son éticos si tienen una probabilidad razonable de generar la respuesta correcta a la pregunta de investigación, y los estudios aleatorizados tienen más posibilidad de llevar a un resultado correcto y concluyente que los diseños no aleatorizados. Además, la base ética de cualquier ensayo es la incertidumbre de si la intervención será beneficiosa o nociva. Esta incertidumbre, denominada equilibrio, significa que la elección de intervenciones basada en pruebas no es posible y justifica la asignación aleatoria.

Diseños intragrupo

Los diseños que no incluyen un grupo testigo separado pueden ser opciones útiles para algunos tipos de preguntas. En un diseño de series cronológicas se realizan medidas antes y después de que cada participante reciba la intervención (fig. 11-3). Por tanto, cada participante actúa como su propio testigo para evaluar el efecto del tratamiento. Esto significa que características individuales como la edad, el sexo y factores genéticos no solo se equilibran (como en los estudios entre grupos), sino que realmente se eliminan como variables de confusión.

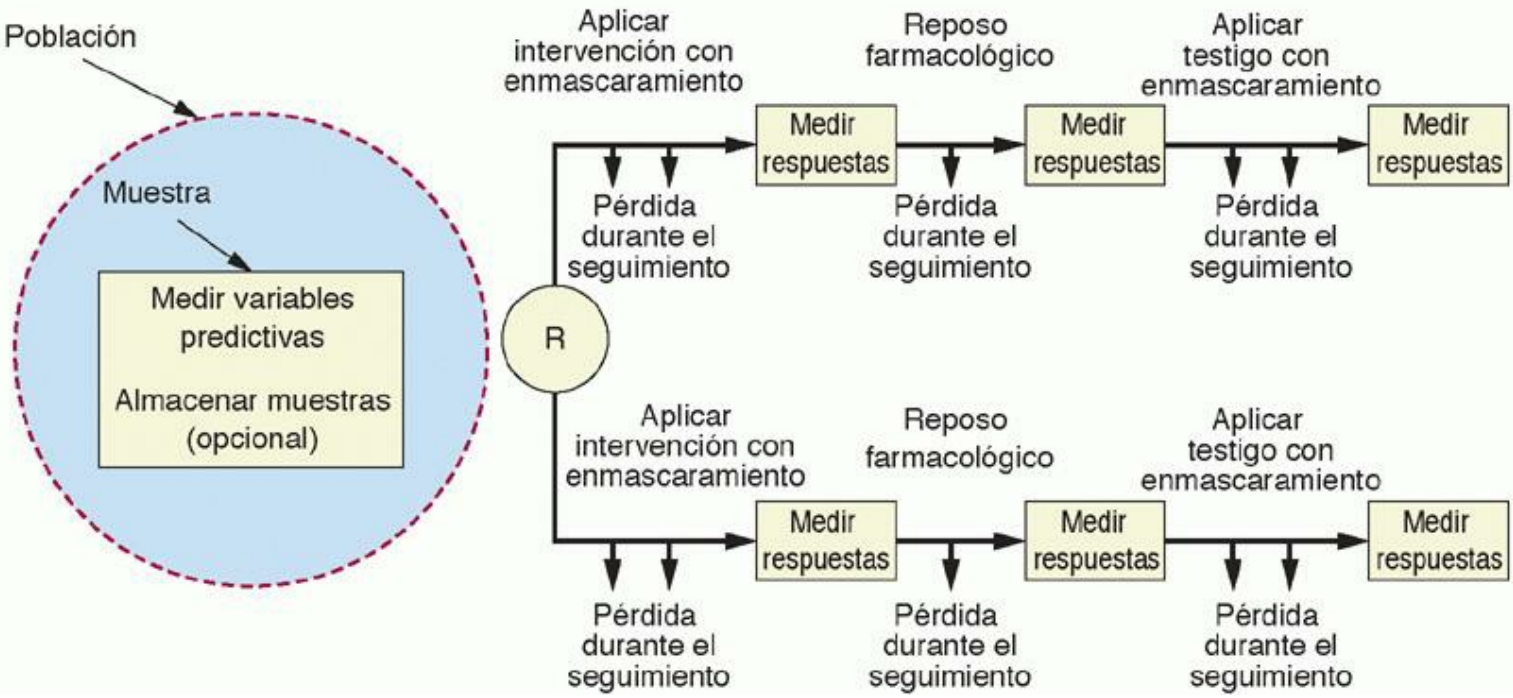


▪FIGURA 11-3. En un estudio de series cronológicas, los pasos son:
• Seleccionar una muestra de participantes de una población idónea para recibir la intervención.
• Medir las variables predictivas y (si procede) el nivel inicial de la variable de respuesta.
• Considerar la opción de almacenar suero, imágenes, etc., para su análisis posterior.
• Aplicar la intervención a toda la cohorte.
• Seguir la cohorte a lo largo del tiempo, minimizando la pérdida durante el seguimiento y evaluando el cumplimiento de la intervención.
• Medir las variables de respuesta.
• Suspender la intervención, continuar el seguimiento y medir de nuevo las variables de respuesta, y posteriormente reiniciar la intervención, y así sucesivamente.
Herramientas de imágenes

El principal inconveniente de los diseños intragrupo es la falta de un grupo testigo *concurrente*. La aparente eficacia de la intervención podría deberse a efectos de aprendizaje (los participantes tienen mejores resultados en las pruebas cognitivas de seguimiento porque aprendieron de la prueba inicial), regresión a la media (se observó que los participantes que fueron seleccionados para el ensayo por tener hipertensión arterial en situación basal tenían una presión arterial inferior durante el seguimiento, sencillamente por una variación aleatoria de la presión arterial) o tendencias a largo plazo (las infecciones respiratorias superiores fueron menos frecuentes durante el seguimiento, porque la temporada de gripe finalizó antes de que hubiera terminado el seguimiento). En los diseños intragrupos a veces se usa una estrategia de iniciar e interrumpir repetidamente el tratamiento. Si el inicio y la interrupción repetidos de la intervención producen patrones similares, es un dato importante que confirma que estos cambios se deban al tratamiento. Este método solo es útil cuando la variable de respuesta responde rápidamente y de modo reversible a la intervención. El diseño tiene una aplicación clínica en los estudios «N de uno», en los que un paciente concreto puede alternar entre las versiones activa e inactiva de un fármaco (usando un placebo de aspecto idéntico preparado por la farmacia local) para detectar su respuesta particular al tratamiento (12).

Diseño cruzado

El diseño cruzado tiene características de los diseños intragrupo y entre grupos (fig. 11-4). Se asigna aleatoriamente a la mitad de los participantes a empezar con el período testigo y luego se los cambia al tratamiento activo; la otra mitad empieza con el tratamiento activo y cambia luego al testigo. Este método permite realizar análisis entre los grupos, así como dentro de ellos. Las ventajas de este diseño son importantes: reduce al mínimo la posibilidad de confusión, porque cada participante actúa como su propio testigo y el análisis emparejado aumenta la potencia estadística del estudio, de modo que se necesitan menos participantes. No obstante, los inconvenientes también son importantes: duración doble del estudio, gastos añadidos necesarios para medir la variable de respuesta al comienzo y al final de cada período de cruzamiento, y complejidad añadida del análisis y la interpretación creados por el problema de posibles efectos residuales. Un efecto de este tipo es la influencia residual de la intervención sobre la respuesta durante el período posterior a la interrupción: por ejemplo, la presión arterial no regresa a los niveles basales en los meses siguientes a un ciclo de tratamiento diurético. Para reducir este efecto, el investigador puede introducir un período sin tratamiento de reposo farmacológico entre tratamientos, con la esperanza de que la variable de respuesta regrese al valor inicial antes de iniciar la siguiente intervención, pero es difícil saber si todos los efectos residuales se han eliminado. En general, los estudios cruzados son una buena elección cuando el número de participantes del estudio es bajo y la respuesta se produce rápidamente y de forma reversible tras la aplicación de la intervención.



▪FIGURA 11-4. En un estudio aleatorizado cruzado, los pasos son:
• Seleccionar una muestra de participantes de una población adecuada para recibir la intervención.
• Medir las variables predictivas y (si procede) el nivel inicial de la variable respuesta.
• Asignar aleatoriamente la intervención y la situación testigo con enmascaramiento.
• Seguir las cohortes a lo largo del tiempo, minimizando la pérdida durante el seguimiento y evaluando el cumplimiento de las situaciones de intervención y testigo.
• Medir las variables de respuesta.
• Suspender las situaciones de intervención y testigo y aplicar un período de reposo farmacológico para reducir los efectos residuales, si procede.
• Aplicar la intervención al anterior grupo testigo y la situación de testigo al anterior grupo de intervención, y medir la respuesta después de seguir las cohortes a lo largo del tiempo.
Herramientas de imágenes

Puede ser adecuada una variación del diseño cruzado cuando no se puede ocultar la asignación de la intervención en estudio y los participantes piensan que la intervención es mucho más deseable que la situación testigo (como una nueva técnica no invasiva). En esta situación, en la que puede ser muy difícil encontrar participantes idóneos que estén dispuestos a ser aleatorizados, un abordaje excelente puede ser la aleatorización a intervención inmediata o a una situación testigo de lista de espera (diferida). Otra situación en la que un testigo de lista de espera puede ser adecuado es cuando una comunidad, un colegio, un gobierno o una entidad similar han decidido que todos los miembros de un grupo deben recibir una intervención a pesar de que hay pocos datos sobre su eficacia. En esa situación, se puede considerar que la aleatorización a no recibir la intervención no es ética, mientras que puede ser aceptable la aleatorización a una intervención diferida.

El diseño de lista de espera da la oportunidad de realizar una comparación aleatorizada entre los grupos de intervención inmediata y testigo de lista de espera. Además, se pueden combinar los dos períodos de intervención (intervención inmediata en un grupo e intervención diferida en el otro) para aumentar la potencia para una comparación intragrupo antes y después de la intervención. Por ejemplo, en un estudio en el que se aleatorizara a mujeres con fibroides sintomáticos a un nuevo tratamiento que es menos invasivo que la histerectomía (embolización de la arteria uterina) en comparación con la lista de espera, las testigos de la lista de espera no recibirían ningún tratamiento durante el período inicial, y posteriormente se les ofrecería embolización de la arteria uterina al comienzo del siguiente período. Posteriormente se pueden combinar las mediciones de los cambios de la puntuación intragrupo de síntomas del fibroide en todas las participantes a las que se aplicó la intervención.

Este diseño tiene la ventaja de hacer que la inclusión de pacientes sea mucho más viable en un estudio en el que la intervención es muy deseable, y de permitir una comparación aleatorizada en situaciones en las que todos los pacientes idóneos finalmente recibirán una intervención. Sin embargo, la respuesta se debe producir en un período de tiempo corto (o el período de espera llegará a ser prohibitivamente largo). Además, aplicar la intervención al grupo testigo al final del estudio prolonga la duración del seguimiento y puede ser costoso.

Ensayos para la aprobación de nuevas intervenciones por las autoridades de registro sanitario

Muchos ensayos se realizan para analizar la eficacia y la seguridad de nuevos tratamientos para los que se pediría la aprobación de comercialización a la U.S. Food and Drug Administration (FDA) o a otros organismos reguladores. Los ensayos también se realizan para determinar si fármacos que están aprobados por la FDA para una afección podrían aprobarse para el tratamiento o la prevención de otras afecciones. El diseño y la realización de estos ensayos suelen ser iguales que los de otros ensayos, aunque deben tenerse en cuenta los requisitos de las agencias de registro sanitario.

La FDA publica directrices generales y específicas sobre cómo deben realizarse esos ensayos (buscar por «FDA» en la web). Puede ser prudente que los investigadores y el personal que realiza ensayos con el fin de obtener la aprobación de la FDA para un nuevo medicamento o dispositivo realicen formación específica en estas directrices generales, denominadas Prácticas clínicas correctas (cap. 17). Además, la FDA proporciona directrices específicas para los estudios de determinadas respuestas. Por ejemplo, los estudios destinados a obtener la aprobación de la FDA para tratamientos para los sofocos en las mujeres menopáusicas deben incluir actualmente participantes con al menos 7 sofocos diarios o 50 a la semana. Las directrices de la FDA se actualizan periódicamente, y otros organismos reguladores internacionales disponen de directrices similares.

Los ensayos para la aprobación por las autoridades de registro de nuevos tratamientos se describen, generalmente, por fases. El sistema se refiere a una progresión ordenada en el estudio de un nuevo tratamiento, desde los experimentos en animales, cultivos celulares o hísticos humanos (preclínicos) y el tratamiento inicial sin enmascaramiento y no controlado de algunos voluntarios humanos para comprobar la seguridad (fase I), hasta ensayos aleatorizados pequeños o series cronológicas que analizan el efecto de diversas dosis sobre los efectos secundarios y los biomarcadores o las respuestas clínicas (fase II) o ensayos aleatorizados lo suficientemente grandes como para verificar la hipótesis de que el tratamiento mejora la afección a tratar (como la hipertensión arterial) o reduce el riesgo de la enfermedad (como la cerebrovascular) con una seguridad aceptable (fase III) (tabla 11-1). La FDA habitualmente define los criterios de valoración para los estudios de fase III que son necesarios para obtener la autorización de comercialización del nuevo fármaco. La fase IV se refiere a grandes estudios, que pueden ser ensayos aleatorizados, aunque a menudo son estudios de observación extensos que se realizan después de haberse aprobado el fármaco. Estos estudios en ocasiones se diseñan para evaluar la incidencia de acontecimientos adversos graves cuando el fármaco se usa en poblaciones de gran tamaño, o para estudiar otros usos del fármaco que pudieran ser aprobados por la FDA. En ocasiones, los estudios de fase IV no tienen un objetivo científico claro, aunque se realizan para presentar los nuevos fármacos a los médicos y los pacientes.

Preclínica	Estudios en cultivos celulares, tejidos y animales
Fase I	Estudios no enmascarados y no controlados en algunos voluntarios, para evaluar la seguridad
Fase II	Ensayos o series cronológicas aleatorizados y relativamente pequeños para evaluar la tolerancia y los efectos de diferentes intensidades o dosis de la intervención sobre biomarcadores o respuestas clínicas
Fase III	Ensayos con enmascaramiento, aleatorizados y relativamente grandes para evaluar el efecto del tratamiento sobre respuestas clínicas y acontecimientos adversos
Fase IV	Grandes ensayos o estudios de observación realizados después de que el tratamiento haya sido aprobado por la FDA para valorar la incidencia de efectos secundarios graves e infrecuentes y evaluar otros usos terapéuticos

Estudios piloto

El éxito en el diseño y la realización de un ensayo clínico exige una amplia información sobre el tipo, la dosis y la duración de las intervenciones, el efecto probable de la intervención sobre la respuesta, los posibles efectos adversos, la viabilidad de reunir, aleatorizar y mantener a los participantes en el ensayo, y probablemente los costes. A menudo, el único modo de obtener parte de esta información es realizar un buen estudio piloto.

Los estudios piloto varían desde una breve prueba de la viabilidad en un pequeño número de participantes hasta un extenso estudio en cientos de participantes (como preparación para una importante inversión en un estudio multicéntrico de varios años de duración). Deben planificarse tan cuidadosamente como el ensayo principal, con objetivos y métodos claros. Muchos estudios piloto se centran fundamentalmente en determinar la viabilidad, el tiempo y el coste necesarios para reunir cantidades adecuadas de participantes

elegibles, y descubrir si están dispuestos a aceptar la aleatorización y cumplir con la intervención. Los estudios piloto también pueden diseñarse para demostrar que las mediciones previstas, los instrumentos de recopilación de datos y los sistemas de gestión de los datos son viables y eficaces. En los estudios piloto que se realizan principalmente para estudiar la viabilidad no suele incluirse un grupo testigo.

Un objetivo importante de muchos estudios piloto es definir la intervención óptima: la frecuencia, la intensidad y la duración de la intervención que producirán la mínima toxicidad y la máxima eficacia.

Los estudios piloto se usan, en ocasiones, para obtener estimaciones de los parámetros necesarios para estimar el tamaño de la muestra. Las estimaciones sólidas de la incidencia de la variable de respuesta o de la medida de respuesta principal en el grupo del placebo, el efecto de la intervención sobre la variable de respuesta principal (magnitud del efecto) y la variabilidad estadística de esta respuesta son fundamentales para la planificación del tamaño de la muestra. En la mayoría de los casos, es mejor obtener estas estimaciones de estudios a escala completa publicados de intervenciones similares en participantes similares. Cuando no se disponga de esos datos, puede ser útil obtener estimaciones en un estudio piloto, aunque el tamaño de la muestra de los estudios piloto habitualmente es tan pequeño que la magnitud del efecto y la varianza calculadas son inestables, con intervalos de confianza muy amplios.

Muchos ensayos no alcanzan la potencia estimada no porque el efecto de la intervención sea menor que el previsto, sino porque la incidencia de episodios de respuesta dicotómica en el grupo del placebo es mucho menor de lo esperado. Esto probablemente se produce porque las personas que cumplen los criterios de inclusión para un ensayo clínico y que aceptan la aleatorización están más sanas que la población general con la afección de interés. Por lo tanto, es esencial determinar la incidencia de la respuesta en el grupo del placebo, lo que se puede realizar evaluando el grupo del placebo de estudios previos con participantes similares, o asignando aleatoriamente a los pacientes al placebo en un estudio piloto.

Un estudio piloto debe tener un protocolo corto pero completo (aprobado por el comité de revisión institucional), formularios para la recopilación de datos y planes de análisis. Las variables deben incluir las típicas determinaciones basales y las variables predictivas y de respuesta incluidas en un estudio competo, y también cálculos del número de personas disponibles o accesibles para su inclusión, el número de las personas con las que se ha establecido contacto o que responden usando diferentes técnicas de inclusión, el número y la proporción idóneos para el ensayo, las personas que son idóneas pero rechazan (o dicen que rechazarían) la aleatorización, el tiempo y el coste de la reunión y la aleatorización, y estimaciones del cumplimiento de la intervención y otros aspectos del protocolo, como las consultas del estudio. Habitualmente es útil pedir tanto a los participantes como al personal que hagan comentarios después del estudio piloto, para que den sus opiniones sobre cómo mejorar los métodos del estudio.

Un buen estudio piloto necesita tiempo suficiente y puede ser costoso, pero mejora notablemente la posibilidad de financiación de importantes ensayos clínicos y la probabilidad de que el ensayo se realice con éxito.

Volver al principio

▪ REALIZACIÓN DE UN ENSAYO CLÍNICO

Seguimiento y cumplimiento del protocolo

Si un número importante de participantes del estudio no recibe la intervención de este, no cumple el protocolo o se pierde durante el seguimiento, es probable que los resultados del ensayo se infravaloren o estén sesgados. En la tabla 11-2 se esbozan las estrategias para aumentar al máximo el seguimiento y el cumplimiento.

El efecto de la intervención (y la potencia del ensayo) se reduce en la medida en que los participantes no la reciben. El investigador debe tratar de escoger para el estudio una intervención o un fármaco que sea fácil de aplicar o tomar, y que se tolere bien. Es probable que el cumplimiento sea bajo si una intervención conductista precisa horas de consulta con los participantes. Los fármacos que pueden tomarse en una sola dosis diaria son más fáciles de recordar y, por tanto, se prefieren. El protocolo debe incluir provisiones que fomentarán el cumplimiento, como enseñar a los participantes a tomar un comprimido en un momento determinado de la rutina matutina, entregar pastilleros marcados con el día de la semana o enviar recordatorios al teléfono móvil.

También se debe tener en cuenta la forma óptima de medir el cumplimiento de la intervención, usando métodos como la descripción por el propio paciente, el conteo de las pastillas, recipientes para las pastillas con chips informáticos que registren cuándo se abre el recipiente, y concentraciones séricas y urinarias de los metabolitos. Esta información puede identificar a los participantes que no cumplen, de modo que puedan instaurarse métodos para mejorar el cumplimiento y el investigador pueda interpretar adecuadamente los hallazgos del estudio.

El cumplimiento de las consultas del estudio y las mediciones puede mejorarse comentando en qué consiste el estudio antes de obtener el

consentimiento, programando las consultas a una hora conveniente y con suficiente personal para evitar la espera, llamando o enviando correos electrónicos al participante el día anterior a cada consulta, y reembolsando los gastos de viaje y aparcamiento y otros gastos extras.

PRINCIPIO	EJEMPLO
Elegir participantes que es probable que cumplan con la intervención y el protocolo	Exigir que se acuda a dos o más consultas antes de la aleatorización Excluir a los que no cumplan un período de preinclusión antes de la aleatorización Excluir a los que es probable que se muden o que no cumplan
Hacer la intervención fácil	Usar un solo comprimido una vez al día, si es posible
Hacer que las consultas del estudio sean oportunas y agradables	Programar las consultas con suficiente frecuencia para mantener un estrecho contacto, pero no tan frecuentes como para cansar Programar las consultas por la tarde o en fin de semana, o recoger información por teléfono o correo electrónico Contar con personal adecuado y bien organizado para evitar esperas Reembolsar los gastos de transporte y aparcamiento Establecer relaciones interpersonales con los participantes
Realizar mediciones del estudio indoloras, útiles e interesantes	Elegir pruebas no cruentas e informativas que, de otro modo, serían costosas o inaccesibles Entregar los resultados de las pruebas de interés a los participantes, y consejos o remisiones adecuados
Animar a los participantes a que continúen en el ensayo	No suspender nunca el seguimiento por violaciones del protocolo, acontecimientos adversos o finalización de la intervención Enviar a los participantes felicitaciones por su cumpleaños y por vacaciones Enviar boletines informativos y mensajes de correo electrónico Hacer hincapié en la importancia científica del cumplimiento y del seguimiento
Encontrar participantes perdidos durante el seguimiento	Buscar contactos de los participantes Usar un servicio de búsqueda

La imposibilidad de seguir a los participantes del ensayo y de medir la respuesta de interés puede producir resultados sesgados, disminución de la credibilidad de los hallazgos y reducción de la potencia estadística. Por ejemplo, en un ensayo de un aerosol nasal de calcitonina para reducir el riesgo de fracturas osteoporóticas se comunicó que el tratamiento redujo el riesgo de fractura en un 36 % (13). Sin embargo, se perdió durante el seguimiento a alrededor del 60 % de los participantes aleatorizados, y no se supo si se habían producido fracturas en ellos. Como el número total de fracturas era pequeño, incluso algunas fracturas en los participantes a los que se perdió durante el seguimiento podrían haber alterado los hallazgos del ensayo, cuya credibilidad disminuyó a causa de esta incertidumbre

(14).

Incluso si los participantes generan una violación del protocolo o interrumpen la intervención del ensayo, debe realizarse el seguimiento de modo que sus respuestas puedan usarse en el análisis de intención de tratar (véase «Análisis de los resultados», en este capítulo). En muchos ensayos se suspende el seguimiento de los participantes que violan el protocolo participando en otro ensayo, no acudiendo a las consultas del estudio o interrumpiendo su intervención; esto puede producir resultados sesgados o imposibles de interpretar. Considere, por ejemplo, un fármaco que causa un efecto secundario sintomático que lleva a una interrupción más frecuente de la medicación del estudio en los que reciben el tratamiento activo que en los que toman el placebo. Si no se mantiene el seguimiento de los participantes que interrumpen la medicación del estudio, se puede introducir un sesgo en los hallazgos en caso de que el efecto secundario se asocie a la respuesta principal o a un acontecimiento adverso grave (AAG).

Algunas de las estrategias para lograr un seguimiento completo son similares a las comentadas en los estudios de cohortes (cap. 7). Al principio del estudio debe informarse a los pacientes de la importancia del seguimiento, y los investigadores deberán registrar el nombre, la dirección, la dirección de correo electrónico y el número de teléfono de uno o dos familiares o conocidos que siempre sepan dónde está el participante. Además de aumentar la capacidad del investigador para evaluar el estado vital, la posibilidad de contactar con los participantes por teléfono o correo electrónico puede permitir el acceso a medidas de respuestas indirectas de los que rechazan acudir a una consulta al final. En el ensayo clínico *The Heart and Estrogen/Progestin Replacement Study* se usaron todas estas estrategias: el 89 % de las mujeres acudió a la última consulta clínica tras un promedio de 4 años de seguimiento, otro 8 % tuvo un último contacto telefónico para verificar la respuesta, y se determinó la información sobre el estado vital de cada una de las participantes restantes utilizando cartas certificadas, contactos con familiares próximos y un servicio de rastreo (15).

El diseño del ensayo debe facilitar al máximo posible que los participantes acepten la intervención y realicen las visitas de seguimiento y las mediciones. Si las consultas son largas y estresantes, es posible que algunos participantes no acudan. Es más probable que regresen a consultas que conlleven pruebas no cruentas, como tomografía computarizada, que a las que sí suponen pruebas cruentas, como angiografía coronaria. Recoger información por teléfono o con medios electrónicos puede mejorar el cumplimiento de los pacientes para los que las consultas son difíciles. Por otro lado, los participantes pueden perder el interés en un ensayo si no tienen ninguna recompensa social o interpersonal por su participación. Pueden cansarse de consultas que se programan mensualmente, y pueden perder el interés si se programan anualmente. El seguimiento mejora haciendo que la experiencia en el ensayo sea positiva y agradable para los participantes: diseñar medidas y procedimientos indoloros e interesantes; realizar pruebas de las que de otro modo no dispondrían; proporcionar los resultados de las pruebas a los participantes (salvo que se trate de pruebas de investigación especializadas que todavía no se utilicen en la práctica clínica); enviar boletines, mensajes de texto o notas por correo electrónico; abrir páginas en redes sociales; enviar tarjetas para felicitar por las vacaciones y el cumpleaños; entregar regalos baratos; y desarrollar fuertes relaciones interpersonales con un equipo entusiasta y amable.

Dos aspectos del diseño que son específicos de los ensayos clínicos pueden mejorar el cumplimiento y el seguimiento: las consultas de selección antes de la aleatorización y un período de preinclusión. Pedir a los participantes que acudan a una o dos consultas de selección antes de la aleatorización puede descartar ya a los participantes que no pueden acudir a estas consultas. El truco está en poner trabas para entrar en el ensayo en un nivel lo suficientemente elevado como para excluir a los que no cumplirán después, pero no tan elevado como para excluir a participantes que tendrán un cumplimiento satisfactorio.

Puede que sea útil contar con un período de preinclusión para aumentar la proporción de participantes del estudio que cumplen la intervención y los procedimientos de seguimiento. Durante el período inicial, todos los participantes reciben un placebo. Un determinado tiempo después (generalmente unas semanas) solo se aleatorizará a los que hayan cumplido con la intervención (p. ej., tomar al menos el 80 % del placebo asignado). Excluir así a los participantes no cumplidores antes de realizar la aleatorización puede aumentar la potencia del estudio y permitir un cálculo mejor de los efectos completos de la intervención. Sin embargo, un período de preinclusión retrasa la inclusión en el estudio, la proporción de participantes excluidos es generalmente pequeña y los participantes aleatorizados a recibir el fármaco activo pueden apreciar un cambio en su medicación tras la aleatorización, lo que contribuye al desenmascaramiento. Tampoco está claro que una preinclusión con placebo sea más eficaz para aumentar el cumplimiento que el requisito de que los participantes acudan a una o más consultas antes de la aleatorización. Si no hay una razón específica para sospechar que el cumplimiento del estudio va a ser malo, probablemente no sea necesario incluir un período de preinclusión en su diseño.

Una variante del diseño con preinclusión con placebo es el uso del fármaco activo en lugar del placebo en ese período. Además de aumentar el cumplimiento de los participantes, un período de preinclusión con el fármaco activo puede seleccionar a los participantes que toleran la intervención y responden a ella; se puede utilizar como criterio de aleatorización la ausencia de efectos adversos o la presencia de un efecto deseado del tratamiento sobre un biomarcador asociado a la variable de respuesta. Por ejemplo, en un estudio controlado con placebo en el que se evaluó el efecto de la nitroglicerina sobre la masa ósea, los investigadores utilizaron un período de preinclusión con el fármaco activo de 1 semana y excluyeron a las mujeres que interrumpieron la nitroglicerina por cefalea (16). Este diseño incrementó al máximo la potencia, porque aumentó la proporción del grupo asignado a la intervención que toleró el fármaco y tenía probabilidad de cumplir el tratamiento. Sin embargo, los hallazgos de los ensayos que usan estas estrategias no pueden generalizarse a los

pacientes excluidos.

El uso de una preinclusión con el fármaco activo también puede causar una infravaloración de la incidencia de efectos adversos. En un ensayo del efecto del carvedilol sobre la mortalidad en 1904 pacientes con insuficiencia cardíaca congestiva, se usó un período de preinclusión con el fármaco activo de 2 semanas. Durante la preinclusión, 17 personas empeoraron de su afección y 7 fallecieron (17). No se aleatorizó a estas personas en el ensayo, y estos efectos adversos del tratamiento farmacológico no se incluyeron como resultados.

Evaluación y adjudicación de las respuestas

Los datos para considerar que se ha producido una respuesta pueden proceder de muchas fuentes: descripción por los propios pacientes, cuestionarios estandarizados, registros administrativos o clínicos, pruebas de laboratorio o estudios de imagen, mediciones especiales, etc. La mayor parte de las respuestas comunicadas, como el antecedente de accidente cerebrovascular o la comunicación de un participante de que ha dejado de fumar, no son exactas al 100 %. Las respuestas referidas por el propio paciente que son importantes para el ensayo deben confirmarse cuando sea posible. La incidencia de una enfermedad, como el accidente cerebrovascular, suele juzgarse mediante:

- La creación de criterios claros para la respuesta (p. ej., un nuevo déficit neurológico persistente con lesión correspondiente en la tomografía computarizada o la resonancia magnética).
- Recopilando los documentos médicos necesarios para realizar la evaluación (p. ej., resúmenes del alta e informes radiológicos).
- Con expertos que desconozcan la asignación y que revisen cada posible caso y opinen si se han cumplido los criterios para el diagnóstico.

La adjudicación la realizan a menudo dos expertos que trabajan por separado, y luego resuelven los casos discordantes comentándolos entre ellos o con un tercer experto. Sin embargo, la implicación de múltiples expertos para la adjudicación puede ser costosa, y para respuestas sencillas en estudios de pequeño tamaño puede ser suficientemente exacto que haya un único investigador que realice la adjudicación. Lo importante es que cualquier persona que participe en la recogida de la información y la adjudicación debe desconocer la asignación del tratamiento.

Supervisión de los ensayos clínicos

Los investigadores deben asegurarse de que los participantes no estén expuestos a una intervención perjudicial, de que no se les niegue una intervención beneficiosa y de que no continúen en un ensayo si es probable que no se obtenga respuesta a la pregunta de la investigación. Se deben seguir cada una de estas tres consideraciones durante el transcurso del estudio para ver si se debe finalizar precoz mente el estudio:

- Finalización por perjuicios. La razón de más peso para hacer un seguimiento de los ensayos clínicos es asegurarse de que la intervención no se volverá perjudicial inesperadamente. Si se piensa que va a existir claramente un perjuicio y que va a superar a los beneficios, deberá interrumpirse el ensayo.
- Finalización por efecto beneficioso. Si una intervención es más eficaz de lo que se calculó cuando se diseñó el ensayo, puede observarse la mejoría estadísticamente significativa al principio del mismo. Cuando se ha demostrado un claro efecto beneficioso, puede que no sea ético seguir el ensayo y retrasar el ofrecimiento de la intervención a los participantes con placebo y a otros pacientes que pudieran beneficiarse.
- Finalización por futilidad. Si hay muy pocas probabilidades de responder a la pregunta de la investigación, puede no ser ético seguir con un ensayo que requiere tiempo y esfuerzo, y que puede causar molestias y riesgo a los participantes. Si un ensayo clínico está programado para continuar durante 5 años, por ejemplo, pero a los 4 hay pocas diferencias en la incidencia de respuestas en los grupos de intervención y testigo, la «potencia condicional» (la probabilidad de rechazar la hipótesis nula en el tiempo restante, a la vista de los resultados obtenidos hasta ese momento) se vuelve muy pequeña, y hay que pensar en interrumpir el ensayo. A veces, los ensayos se interrumpen pronto, si los investigadores no son capaces de reunir o conservar suficientes participantes para tener una potencia adecuada para responder a la pregunta de la investigación, o si el cumplimiento de la intervención es muy escaso.

Podría ocurrir que la pregunta de la investigación se respondiera en otros ensayos antes de que acabase uno determinado. Es deseable contar con más de un ensayo que proporcione datos sobre una determinada pregunta, pero, si se dispone de pruebas definitivas de un efecto beneficioso o de un efecto perjudicial en uno, puede no ser ético continuar el estudio.

La mayor parte de los ensayos clínicos incluyen un plan de seguimiento intermedio. Los ensayos financiados por los *National Institutes of Health* (NIH) precisan, generalmente, un seguimiento intermedio, incluso si se considera que la intervención es segura (como una

intervención conductista para reducir peso). En la planificación de cada ensayo clínico deberá considerarse el modo en que deberá realizarse el seguimiento. En ensayos pequeños con intervenciones probablemente inocuas, los investigadores podrían controlar la seguridad o citar un solo control de seguridad y datos independiente. En los grandes ensayos, y en los ensayos en los que los efectos adversos de la intervención se desconocen o pueden ser peligrosos, el seguimiento intermedio suele realizarlo un comité, generalmente conocido como comité de supervisión de los datos y la seguridad (CSDS), que está formado por especialistas en la enfermedad o afección que se estudia, bioestadísticos, expertos en ensayos clínicos, expertos en ética y, a veces, un representante del grupo de pacientes que se está estudiando. Estos expertos no intervienen en el ensayo y no deben tener ningún interés personal ni económico en su continuación. Las directrices y procedimientos del CSDS deben detallarse por escrito antes de empezar el estudio. La guía para desarrollar los procedimientos de los CSDS la proporcionan la FDA y los NIH. En la tabla 11-3 se muestran los temas que hay que incluir en esas directrices.

La interrupción de un ensayo clínico debe ser siempre una decisión cuidadosa que equilibre la responsabilidad ética frente a los participantes y el progreso del conocimiento científico. Siempre que se interrumpa pronto un ensayo clínico se perderá la posibilidad de obtener resultados más concluyentes. A menudo, la decisión es compleja, y deben sopesarse los posibles riesgos para los participantes frente a los posibles beneficios. Las pruebas estadísticas de significación utilizando uno de los métodos que compensan los múltiples análisis de los hallazgos (apéndice 11B) proporcionan información importante, aunque no concluyente, para la interrupción de un estudio. Debe evaluarse la constancia de las tendencias a lo largo del tiempo y los efectos sobre las respuestas relacionadas, y deberán considerarse cuidadosamente las consecuencias que tendrá la interrupción temprana del estudio sobre la credibilidad de los hallazgos (ejemplo 11-2).

Hay muchos métodos estadísticos para supervisar los resultados provisionales de un ensayo. El análisis repetido de los resultados de un estudio («múltiples vistazos») es una forma de comprobar hipótesis múltiples y aumenta la probabilidad de que se produzca un error de tipo I. Por ejemplo, si se usa $\alpha = 0{,}05$ para cada prueba provisional, y se analizan los resultados de un ensayo cuatro veces durante su realización y de nuevo al final, la probabilidad de cometer un error de tipo I aumenta desde el 5 % a alrededor del 14 % (18). Para abordar este problema, los métodos estadísticos para la supervisión provisional suelen disminuir el valor α de cada prueba, de modo que el valor de α total sea próximo a 0,05. Hay múltiples métodos para decidir cómo «gastar α» (v. apéndice 11B).

Elementos que se deben seguir
Inclusión
Aleatorización
Cumplimiento de la intervención y el enmascaramiento
Cumplimiento completo del seguimiento
Variables importantes
De respuesta
Efectos adversos
Posibles cointervenciones
Quién hará el seguimiento
El investigador del ensayo o un único supervisor si el ensayo es pequeño y tiene pocos riesgos

En los demás casos un comité de supervisión de los datos y de la seguridad independiente

Métodos para el seguimiento provisional

Especificar por adelantado el abordaje estadístico y la frecuencia del seguimiento

Importancia del juicio y el contexto, además de las reglas estadísticas de finalización

Cambios del protocolo que pueden surgir del seguimiento

Finalizar el ensayo

Modificar el ensayo

Poner fin a un grupo del ensayo

Añadir nuevas mediciones necesarias para seguir la seguridad

Suspender en participantes de riesgo elevado

Ampliar la duración del ensayo

Aumentar la muestra del ensayo

Analizar los resultados: intención de tratar y según protocolo

El análisis estadístico de la hipótesis principal de un ensayo clínico suele ser sencillo. Si la respuesta es dicotómica, el método más sencillo será comparar las proporciones en los grupos del estudio mediante una prueba de χ^2. Cuando la respuesta es continua, puede usarse una prueba de la*t*, o una alternativa no paramétrica si la respuesta no está distribuida normalmente. En muchos ensayos clínicos, la duración del seguimiento es diferente para cada participante, por lo que se necesita el uso de métodos de tiempo de supervivencia. Modelos estadísticos más sofisticados, como el análisis de riesgos proporcionales de Cox, pueden lograrlo y, al mismo tiempo, ajustar las distribuciones heterogéneas y casuales de las variables de confusión basales (19).

Un aspecto importante que se debe considerar en el análisis de los resultados en los ensayos clínicos es la primacía del abordaje analítico de intención de tratar para manejar los «cruzamientos», participantes asignados al grupo de tratamiento activo que no cumplen el tratamiento o que lo interrumpen, y participantes asignados al grupo testigo que terminan recibiendo el tratamiento activo. Un análisis de intención de tratar compara las respuestas entre los grupos del estudio analizando cada participante según su asignación aleatorizada al grupo, independientemente de si cumplió la intervención asignada. Los análisis de intención de tratar pueden infravalorar el efecto completo del tratamiento, pero se protegen frente a fuentes más importantes de resultados sesgados.

Una alternativa al método de intención de tratar es realizar un análisis «según protocolo» en el que solo se incluya a los participantes que hayan cumplido el protocolo. Esto se define de varios modos, pero a menudo solo se incluyen los participantes de ambos grupos que cumplieron con la medicación asignada en el estudio, realizaron una determinada proporción de consultas o mediciones, y no incumplieron de ninguna otra forma el protocolo. Un subgrupo del análisis según protocolo es el análisis «cómo se les trató», en el que solo se incluyen los pacientes que cumplieron la intervención asignada. Estos análisis *parecen* razonables, porque los participantes solo pueden verse

afectados por una intervención que realmente reciben. Sin embargo, los participantes que cumplen el protocolo y el tratamiento del estudio pueden ser diferentes de los que no lo hacen en aspectos que estén relacionados con las respuestas. En el ensayo *Postmenopausal Estrogen-Progestin Interventions* (PEPI) se asignó aleatoriamente a 875 mujeres posmenopáusicas a recibir cuatro regímenes diferentes de estrógenos o estrógenos más progestágenos y placebo (20). De las mujeres asignadas al grupo de estrógenos sin oposición, el 30 % interrumpieron el tratamiento a los 3 años por hiperplasia endometrial, un precursor del cáncer endometrial. Si se eliminara a estas mujeres en un análisis según protocolo, se pasaría por alto la asociación entre el tratamiento estrogénico y el cáncer endometrial.

El principal inconveniente del método de intención de tratar es que se incluirán en el cálculo de los efectos de una intervención asignada a los participantes que elijan no cumplir esa intervención. Por lo tanto, una elevada incidencia de suspensión del tratamiento o de cruce entre tratamientos hará que los análisis de intención de tratar infravaloren la magnitud del efecto del tratamiento. Por este motivo, los resultados de los ensayos se evalúan, a menudo, mediante ambos análisis: de intención de tratar y según protocolo. Por ejemplo, en el ensayo aleatorizado *Women's Health Initiative* del efecto de estrógenos más progestágenos sobre el riesgo de sufrir cáncer de mama, la proporción de riesgo fue de 1,24 ($p = 0,003$) en el análisis por intención de tratar, y de 1,49 ($p < 0,001$) en el análisis de «como se las trató» (21). Si los resultados de los análisis según protocolo y de intención de tratar difieren, los resultados de intención de tratar predominan para cálculos de eficacia, porque conservan el valor de la aleatorización y, a diferencia de los análisis según protocolo, solo pueden sesgar el efecto calculado en la dirección conservadora (favoreciendo la hipótesis nula). No obstante, para cálculos del perjuicio (p. ej., los hallazgos de cáncer de mama), los análisis según se ha tratado a los pacientes o según protocolo proporcionan los cálculos más conservadores, ya que solo puede esperarse que las intervenciones causen perjuicio a las personas expuestas.

Los resultados solo pueden analizarse mediante intención de tratar si se realizan las mediciones de seguimiento, independientemente de si los participantes cumplen o no el tratamiento. Por lo tanto, este siempre debe ser el objetivo.

Análisis de subgrupos

Los análisis de subgrupos se definen como comparaciones entre grupos aleatorizados en un subgrupo de la cohorte del ensayo. El principal motivo de realizar estos análisis es descubrir la modificación del efecto («interacción») en distintos subgrupos, por ejemplo, si el efecto de un tratamiento es diferente en hombres y mujeres. Estos análisis tienen una reputación mixta, porque es fácil que se utilicen mal y pueden llevar a conclusiones erróneas. Con cuidado, sin embargo, pueden proporcionar información auxiliar útil y ampliar las inferencias que se pueden extraer de un ensayo clínico. Para conservar el valor de la aleatorización deben definirse los subgrupos mediante determinaciones realizadas antes de ella. Por ejemplo, en un estudio de denosumab para la prevención de las fracturas se encontró que el fármaco reducía el riesgo de fracturas no vertebrales en un 20 % en mujeres con densidad ósea baja. En los análisis de subgrupos planificados previamente se vio que el tratamiento era eficaz (reducción del riesgo de fractura del 35 %; $p < ,01$) en mujeres con densidad ósea baja en situación inicial, y que el tratamiento no era eficaz en mujeres con mayor densidad ósea en situación inicial ($p = 0,02$ para la modificación del efecto) (22). Es importante señalar que se conserva el valor de la aleatorización: la incidencia de fracturas entre las mujeres aleatorizadas a recibir denosumab se comparó con la incidencia en las mujeres aleatorizadas a recibir el placebo en cada uno de los subgrupos. Los análisis de subgrupos basados en factores posteriores a la aleatorización, como el cumplimiento del tratamiento asignado aleatoriamente, no conservan la utilidad de la aleatorización y muchas veces producen resultados engañosos.

Los análisis de subgrupos pueden producir resultados engañosos por varios motivos. Como son más pequeños que toda la población del ensayo, pueden no tener la suficiente potencia para encontrar diferencias importantes; los investigadores deben evitar afirmar que un fármaco «no fue eficaz» en un subgrupo cuando el hallazgo podría reflejar una potencia insuficiente para encontrar un efecto. A menudo, los investigadores examinan resultados en gran cantidad de subgrupos, lo que aumenta la probabilidad de encontrar un efecto diferente de la intervención en un subgrupo por el efecto del azar. Por ejemplo, si se examinan 20 subgrupos, se podría esperar que se produjeran por azar diferencias en un subgrupo con $p < 0,05$. Para abordar este problema, los análisis planificados en subgrupos deben definirse antes de iniciar el ensayo, y el número de subgrupos analizados debe comunicarse con los resultados del estudio (23). Las afirmaciones sobre las diferentes respuestas en distintos subgrupos deben estar respaldadas por la presencia de una interacción estadísticamente significativa entre el efecto del tratamiento y la característica del subgrupo, y se debe confirmar en un estudio separado la modificación del efecto antes de que se considere que existe realmente.

Volver al principio

▪ RESUMEN

- Hay algunas variaciones en el diseño de ensayos aleatorizados que pueden aumentar notablemente la eficacia en las circunstancias adecuadas:

    - El diseño factorial permite realizar dos ensayos independientes por el precio de uno.
    - La aleatorización por conglomerados permite realizar estudios eficaces de grupos que se producen de forma natural.
    - Los ensayos de ausencia de inferioridad o de equivalencia comparan una nueva intervención con un «tratamiento estándar» ya existente.
    - Los diseños adaptativos incrementan la eficiencia, al permitir que se realicen cambios del diseño basados en análisis provisionales, por ejemplo, alterando la dosis del fármaco en estudio, el número de participantes o la duración del seguimiento.
- También hay otros diseños de estudios clínicos útiles:
    - En los diseños de series cronológicas hay un único grupo en el que se comparan los resultados para cada participante durante los períodos en los que recibe y no recibe la intervención.
    - Los diseños cruzados combinan diseños entre grupos e intragrupo para mejorar el control de la confusión (si los efectos residuales no son problemáticos) y para minimizar el tamaño de la muestra.
- Los estudios para la autorización de nuevos fármacos por las autoridades de registro sanitario se clasifican como:
    - Fase I: estudios pequeños para explorar la dosis y la seguridad.
    - Fase II: estudios aleatorizados de tamaño medio o series cronológicas para estudiar los efectos del fármaco en diferentes dosis.
    - Fase III: estudios aleatorizados extensos para demostrar que los beneficios superan a los perjuicios como base para la autorización por la FDA.
    - Fase IV: estudios de observación poscomercialización extensos para confirmar los efectos beneficiosos y detectar efectos adversos infrecuentes.
- Los estudios piloto son pasos importantes para determinar la aceptabilidad de las intervenciones y la viabilidad, el tamaño, el coste y la duración de los estudios previstos.
- Cuando se realiza un estudio, si un número elevado de participantes en el estudio no cumplen la intervención del estudio o se pierden durante el seguimiento, es probable que los resultados del estudio tengan poca potencia, tengan sesgo o no se puedan interpretar.
- Durante un estudio es necesario el seguimiento provisional por un comité de supervisión de los datos y la seguridad (CSDS) independiente para garantizar la calidad del estudio y para decidir si el estudio debe finalizar precozmente por datos de perjuicio, beneficio o futilidad.
- El análisis de intención de tratar se beneficia del control de la confusión que proporciona la aleatorización, y debe ser el principal método de análisis para evaluar la eficacia. Los análisis según protocolo, un método secundario que proporciona un cálculo de la magnitud del efecto en los participantes que cumplen (interpretado con precaución), es el análisis más conservador de los efectos perjudiciales del tratamiento.
- Los análisis de subgrupos permiten detectar si el efecto del tratamiento está modificado por otras variables; para reducir al mínimo las interpretaciones erróneas, el investigador debe especificar por adelantado los subgrupos, evaluar la significación estadística de las modificaciones del efecto (interacciones) y comunicar el número de subgrupos examinados.

Volver al principio

APÉNDICE 11AEspecificación del límite de ausencia de inferioridad en un estudio de ausencia de inferioridad
NA

Uno de los aspectos más difíciles del diseño de un estudio de ausencia de inferioridad es establecer la pérdida de eficacia de nuevo tratamiento que sería inaceptable (7), a la que se denomina «Δ» y que a menudo se llama límite de ausencia de inferioridad. Esta decisión se basa en consideraciones estadísticas y clínicas de la posible eficacia y las posibles ventajas del nuevo tratamiento, y hace falta un

juicio experto. Se presenta un ejemplo del proceso.

EJEMPLO 11-1. Diseño de un estudio de un nuevo fármaco en comparación con warfarina en pacientes con fibrilación auricular

La warfarina reduce el riesgo de accidente cerebrovascular en pacientes de riesgo elevado con fibrilación auricular, por lo que cualquier nuevo fármaco se debe comparar con este tratamiento estándar. Cuando se utiliza warfarina para reducir el riesgo de accidente cerebrovascular en esa situación, es difícil administrar la dosis correctamente, hacen falta análisis de sangre frecuentes para seguir el nivel de anticoagulación y pueden producirse hemorragias graves. Si se dispusiera de un nuevo fármaco que no tuviera estos inconvenientes, sería razonable preferir este fármaco a la warfarina, aunque su eficacia en la reducción del riesgo de accidente cerebrovascular fuera ligeramente menor.

Un abordaje para establecer el valor de Δ es realizar un metaanálisis de estudios previos de warfarina en comparación con el placebo, y establecer Δ en alguna proporción de la distancia entre el valor nulo y el límite inferior del efecto del tratamiento con warfarina. De manera alternativa, como los estudios incluidos en el metaanálisis muchas veces tienen una calidad variable, puede ser mejor basar Δ en los resultados del estudio aleatorizado de warfarina de mejor calidad que tenga criterios de inclusión, dosis de warfarina y criterios de valoración similares. Es importante definir Δ de manera que haya una elevada probabilidad, teniendo en consideración todos los efectos beneficiosos y perjudiciales, de que el nuevo tratamiento sea mejor que el placebo (6, 7).

Suponga que en un metaanálisis de estudios de buena calidad de warfarina en comparación con el placebo se ve que el tratamiento con warfarina reduce la incidencia de accidente cerebrovascular en pacientes de riesgo elevado con fibrilación auricular desde el 10 % al año hasta aproximadamente el 5 % al año (efecto absoluto del tratamiento = 5 %; IC del 95 %: 4-6 %). A la vista de las ventajas de nuestro nuevo fármaco, ¿qué pérdida de eficacia es clínicamente inaceptable? ¿Tal vez una eficacia absoluta que sea un 2 % menor que la de la warfarina sería aceptable? En este caso declararíamos que el nuevo tratamiento no es inferior a la warfarina si el límite inferior del intervalo de confianza alrededor de la diferencia de la incidencia de accidente cerebrovascular entre la warfarina y el nuevo tratamiento es menor del 2 % (v. fig. 11-2). En un estudio de ausencia de inferioridad también es posible que se encuentre que el nuevo tratamiento es superior al tratamiento ya establecido (ejemplo superior de la fig. 11-2).

APÉNDICE 11BSeguimiento provisional de las respuestas del ensayo y finalización temprana
NA

El seguimiento provisional de los resultados del ensayo para decidir si se debe poner fin a un estudio es una forma de estudio de hipótesis múltiples y, por lo tanto, aumenta la probabilidad de error de tipo I. Para abordar este problema, suele disminuirse α para cada prueba (α_i), de modo que α total sea aproximadamente = 0,05. Hay muchos métodos estadísticos para disminuir α_i.

Uno de los más fáciles de entender es el método de Bonferroni, donde $\alpha_i = \alpha/N$, si N es el número total de pruebas realizadas. Por ejemplo, si α total es 0,05 y se realizan cinco pruebas, α_i para cada prueba es 0,01. Sin embargo, este método tiene dos inconvenientes: precisa el uso de un mismo umbral para interrumpir el ensayo en cualquier análisis provisional, y produce un valor de α muy bajo para el análisis final. La mayoría de los investigadores usarían, en su lugar, un umbral más estricto para detener un ensayo antes en lugar de más tarde, y usarían un α cercano a 0,05 para el análisis final. Además, este método es demasiado conservador, porque supone que cada prueba es independiente. Los análisis provisionales no son independientes, porque cada análisis sucesivo se basa en datos acumulados, algunos de los cuales se incluyeron en análisis anteriores. Por estos motivos, no suele usarse el método de Bonferroni.

Un método usado habitualmente, propuesto por O'Brien y Fleming (24), usa un valor de α_i muy pequeño para la prueba de hipótesis inicial, que aumenta gradualmente para cada prueba, de forma que para la prueba final α_i está próxima a α total. O'Brien y Fleming proporcionaron métodos para calcular α_i si el investigador elige el número de pruebas a realizar y el valor α total. En cada prueba, $Z_i = Z^* (N_i)^{1/2}$, donde Z_i = valor de Z para la *i*ésima prueba; Z^* se determina para lograr el nivel de significación total; y N es el número de pruebas planificadas. Por ejemplo, para cinco pruebas y α total = 0,05, $Z^* = 2,04$; el valor de α inicial es 0,00 001 y α_5 final = 0,046. No es probable que este método lleve a la finalización muy temprana de un ensayo, salvo que exista una notable diferencia en las respuestas entre los grupos aleatorizados. Además, este método evita la incómoda situación de llegar al final de un ensayo y aceptar la hipótesis nula cuando el valor de p es 0,04 o 0,03, aunque α_i para la prueba final disminuya hasta un valor de 0,01.

Un inconveniente importante del método de O'Brien y Fleming es que el número de pruebas y la proporción de datos que se van a estudiar deben decidirse antes de empezar el estudio. DeMets y Lan (25) desarrollaron un método usando una función de gasto de α especificada que proporciona límites de detención continuos. El valor de α_i en un momento dado (o tras una determinada proporción de respuestas) se determina con la función y el número de «vistazos» previos. Con este método no es necesario que se especifique antes del estudio ni el número de «vistazos» ni la proporción de datos que se van a analizar en cada «vistazo». Por supuesto, para cada análisis provisional no planificado adicional realizado, el valor de α total final es algo menor.

Un grupo diferente de métodos estadísticos basados en técnicas de muestreo reducidas sugiere la finalización de un ensayo si no es probable que los futuros datos cambien la conclusión. El problema de las pruebas múltiples es irrelevante, porque la decisión solo se basa en la estimación de lo que los datos mostrarán al final del estudio. Un método habitual es calcular la probabilidad de rechazar la hipótesis nula al final del ensayo, condicionada por los datos acumulados. Típicamente se calculan unos valores de potencia condicional, suponiendo primero que H_0 es verdad (es decir, que cualquier respuesta futura en los grupos de control y con tratamiento se distribuirá por igual) y también que H_a es verdad (es decir, que las respuestas no se distribuirán por igual en los grupos de control y de tratamiento, como se especifica por H_a). También pueden usarse otros cálculos para obtener un espectro completo de magnitudes del efecto razonables. Si la potencia condicional para rechazar la hipótesis nula en todo el espectro de supuestos es baja, no es probable que se rechace la hipótesis nula, y el ensayo podría interrumpirse.

En el ejemplo 11-2 se muestran ejemplos de dos estudios a los que se puso fin precozmente.

EJEMPLO 11-2. Dos estudios que se finalizaron precozmente

Cardiac Arrhythmia Suppression Trial(CAST) (26). La aparición de extrasístoles en supervivientes a un infarto de miocardio (IM) es un factor de riesgo de muerte súbita. En el estudio CAST se evaluó el efecto del tratamiento antiarrítmico (encainida, flecainida o moricizina) sobre el riesgo de muerte súbita en pacientes con arritmias ventriculares asintomáticas o levemente sintomáticas después de un IM. Durante un promedio de 10 meses de tratamiento, los participantes a los que se trató con el fármaco activo tuvieron mayor mortalidad total (el 7,7 % en comparación con el 3,0 %) y mayor incidencia de muerte por arritmia (el 4,5 % en comparación con el 1,5 %) que los pacientes a los que se asignó al placebo. Estaba previsto que el estudio continuara durante 5 años, aunque esta diferencia, grande y con elevada significación estadística, llevó a la interrupción del estudio después de 18 meses.

Physicians' Health Study(27). El *Physicians' Health Study* fue un estudio aleatorizado en el que se evaluó el efecto del ácido acetilsalicílico (325 mg a días alternos) sobre la mortalidad cardiovascular. El estudio se interrumpió después de 4,8 años de un seguimiento previsto de 8 años. Hubo una reducción estadísticamente significativa del riesgo de infarto de miocardio mortal en el grupo tratado (riesgo relativo = 0,56), pero sin diferencias en el número de muertes por enfermedad cardiovascular. La incidencia de muerte por enfermedad cardiovascular observada en el estudio fue mucho menor de lo esperado (88 después de 4,8 años y seguimiento, en contraposición con una cifra esperada de 733), y el estudio se interrumpió debido al efecto beneficioso del ácido acetilsalicílico sobre el riesgo de IM no mortal, asociado a una potencia condicional muy baja para detectar un efecto favorable sobre la mortalidad cardiovascular.

BIBLIOGRAFÍA

1. Ridker PM, Cook NR, Lee I, et al. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. N Engl J Med 2005;352:1293-1304. Citado aquí

2. The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. Control Clin Trials 1998;19:61-109. Citado aquí

3. Walsh M, Hilton J, Masouredis C, et al. Smokeless tobacco cessation intervention for college athletes: results after 1 year. Am J Public Health 1999;89:228-234. Citado aquí

4. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. Am J Epidemiol 1981;114:906-914. Citado aquí

5. Piaggio G, Elbourne DR, Altman DG, et al. Reporting of non-inferiority and equivalence randomized trials. An extension of the CONSORT Statement. JAMA 2006;295:1152-1160. Citado aquí

6. Piaggio G, Elbourne DR, Pocock SJ, et al. Reporting of non-inferiority and equivalence randomized trials. An extension of the CONSORT 2010 statement. JAMA 2012;308:2594-2604. Citado aquí

7. Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of non-inferiority trials. Ann Intern Med 2006;145:62-69. Citado aquí

8 D'Agostino RB Sr., Massaro JM, Sullivan LM, et al. Non-inferiority trials: design concepts and issues—the encounters of academic consultants in statistics. Statist Med 2003;22:169-186. Citado aquí

9. Chang M, Chow S, Pong A. Adaptive design in clinical research: issues, opportunities, and recommendations. J Biopharm Stat 2006;16:299-309. Citado aquí

10. Chalmers T, Celano P, Sacks H, et al. Bias in treatment assignment in controlled clinical trials. N Engl J Med 1983;309:1358-1361. Citado aquí

11. Pocock S. Current issues in the design and interpretation of clinical trials. Br Med J 1985;296:39-42. Citado aquí

12. Nickles CJ, Mitchall GK, Delmar CB, et al. An n-of-1 trial service in clinical practice: testing the effectiveness of stimulants for attention-deficit/hyperactivity disorder. Pediatrics 2006;117:2040-2046. Citado aquí

13. Chestnut CH III, Silverman S, Andriano K, et al. A randomized trial of nasal spray salmon calcitonin in postmenopausal women with established osteoporosis: the prevent recurrence of osteoporotic fractures study. Am J Med 2000;109:267-276. Citado aquí

14. Cummings SR, Chapurlat R. What PROOF proves about calcitonin and clinical trials. Am J Med 2000;109:330-331. Citado aquí

15. Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. JAMA 1998;280:605-613. Citado aquí

16. Jamal SA, Hamilton CJ, Eastell RJ, Cummings SR. Effect of nitroglycerin ointment on bone density and strength in postmenopausal women. JAMA 2011;305:800-805. Citado aquí

17. Pfeffer M, Stevenson L. Beta-adrenergic blockers and survival in heart failure. N Engl J Med 1996;334:1396-1397. Citado aquí

18. Armitage P, McPherson C, Rowe B. Repeated significance tests on accumulating data. J R Stat Soc 1969;132A:235-244. Citado aquí

19. Friedman LM, Furberg C, DeMets DL. Fundamentals of clinical trials, 3rd ed. St. Louis, MO: Mosby Year Book, 1996. Citado aquí

20. Writing Group for the PEPI Trial. Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women. JAMA 1995;273:199-208. Citado aquí

21. Writing group for WHI investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women. JAMA 2001;288:321-333. Citado aquí

22. McClung MR, Boonen S, Torring O, et al. Effect of denosumab treatment on the risk of fractures in subgroup of women with postmenopausal osteoporosis. J Bone Mineral Res 2012;27:211-218. Citado aquí

23. Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine—Reporting of subgroup analyses in clinical trials. NEJM 2007;357:2189-2194. Citado aquí

24. O'Brien P, Fleming T. A multiple testing procedure for clinical trials. Biometrics 1979;35:549-556. Citado aquí

25. DeMets D, Lan G. The alpha spending function approach to interim data analyses. Cancer Treat Res 1995;75:1-27. Citado aquí

26. Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. N Engl J Med 1989;321:406-412. Citado aquí

27. Physicians' Health Study Investigations. Findings from the aspirin component of the ongoing Physicians' Health Study. N Engl J Med 1988;318:262-264. Citado aquí

CAPÍTULO 12 Diseño de estudios de pruebas médicas

Thomas B. Newman
Warren S. Browner
Steven R. Cummings
Stephen B. Hulley

Las pruebas médicas, como las que se realizan para detectar un factor de riesgo, diagnosticar una enfermedad o valorar el pronóstico de un paciente, constituyen un tema importante en la investigación clínica. Los diseños de estudios que se comentan en este capítulo pueden usarse cuando se estudia si, y en quién, debe realizarse una prueba concreta.

La mayor parte de los diseños de estudios de pruebas médicas se parecen a los estudios de observación de los capítulos 7 y 8. Existen, no obstante, algunas diferencias entre la mayoría de los estudios de observación y los que se utilizan para evaluar las pruebas médicas. La más importante es que el objetivo de la mayoría de los estudios de observación es identificar asociaciones estadísticamente significativas (cap. 5) que representan relaciones causales (cap. 9). Por el contrario, la demostración de que el resultado de una prueba tiene una asociación estadísticamente significativa con una enfermedad particular no es en modo alguno suficiente para determinar que la prueba tendría utilidad clínica, y en los estudios de pruebas médicas la causalidad es muchas veces irrelevante. Por lo tanto, los valores de la razón de posibilidades y de *p* son consideraciones secundarias en los estudios de pruebas médicas, que, por el contrario, se centran en parámetros *descriptivos* como sensibilidad, especificidad y cocientes de verosimilitudes, con sus correspondientes intervalos de confianza.

▪ DETERMINACIÓN DE LA UTILIDAD DE UNA PRUEBA

Para que una prueba sea útil debe superar una serie de preguntas de dificultad creciente que analizan su reproducibilidad, exactitud, viabilidad y, lo que es más importante, sus efectos sobre las decisiones clínicas y los resultados (tabla 12-1). Las respuestas favorables a todas estas cuestiones son criterios necesarios, aunque no suficientes, para que merezca la pena realizar una prueba. Por ejemplo, si una prueba ofrece resultados muy diferentes dependiendo de quién la realiza o dónde se realiza, es poco probable que sea útil. Si la prueba rara vez proporciona nueva información, es poco probable que afecte a las decisiones clínicas. Incluso si afecta a las decisiones, si estas no mejoran las respuestas clínicas de los pacientes en los que se hizo la prueba con un riesgo y un coste razonables, es probable que siga sin ser útil.

Por supuesto, si el uso de una prueba mejora los resultados de los pacientes estudiados, pueden inferirse respuestas favorables a las otras cuestiones. Sin embargo, los estudios de si realizar una prueba mejora la evolución de los pacientes son los más difíciles de realizar. En cambio, los posibles efectos de una prueba sobre los resultados clínicos suelen inferirse comparando la exactitud, la seguridad y los costes con los de las pruebas ya existentes. Al desarrollar una nueva prueba diagnóstica o pronóstica, puede que valga la pena considerar qué aspectos de la práctica actual son los que más necesitan una mejora. Por ejemplo, ¿son las pruebas actuales poco fiables, caras, peligrosas o difíciles de realizar?

Aspectos generales de los estudios de pruebas médicas

- Espectro de la gravedad de la enfermedad y de los resultados de la prueba. Como el objetivo de la mayor parte de los estudios de pruebas médicas es extraer inferencias sobre poblaciones realizando determinaciones en muestras, el modo de elegir la muestra tiene un efecto importante sobre la validez de las inferencias. Se produce sesgo de espectro cuando el espectro de enfermedad (o de ausencia de enfermedad) en la muestra difiere del de los pacientes a los cuales el investigador desea generalizar los resultados. En las primeras fases del desarrollo de una prueba diagnóstica, puede ser razonable investigar si una prueba permite distinguir entre los pacientes con enfermedad manifiesta en estadios avanzados y los testigos sanos; si la respuesta es no, el investigador puede volver al laboratorio para elaborar una modificación o una prueba diferente. Sin embargo, posteriormente, cuando la pregunta de la investigación aborde la utilidad clínica de la prueba, el espectro de enfermedad y ausencia de enfermedad debe ser representativo de las personas en las que se utilizará la prueba. Por ejemplo, una prueba que se haya desarrollado comparando pacientes con cáncer pancreático sintomáticos con testigos sanos se podría evaluar posteriormente en una muestra más difícil, pero clínicamente realista, como pacientes consecutivos con dolor abdominal o pérdida de peso de causa no determinada.

PREGUNTA	**DISEÑOS POSIBLES**	**ESTADÍSTICOS PARA LOS RESULTADOS***
¿Qué reproducibilidad	Estudios intra- e interobservador, e intralaboratorio y entre	Proporción de acuerdo, κ, coeficiente de variación, media y distribución de diferencias

tiene la prueba?	laboratorios	(evitar el coeficiente de correlación)
¿Qué exactitud tiene la prueba?	Diseños transversales, de casos y testigos o de cohortes, en los que se compara el resultado de una prueba con un método de referencia	Sensibilidad, especificidad, valor predictivo positivo y negativo, curvas de eficacia diagnóstica y cocientes de verosimilitudes
¿Con qué frecuencia los resultados de las pruebas afectan a las decisiones clínicas?	Estudios de rendimiento diagnóstico, estudios de toma de decisiones clínicas antes y después de la prueba	Proporción anormal, proporción con resultados discrepantes, proporción de pruebas que producen cambios en las decisiones clínicas; coste por resultado anormal o por cambio de decisión
¿Cuáles son los costes, riesgos y aceptabilidad de la prueba?	Estudios prospectivos o retrospectivos	Costes medios, proporciones de los que sufren efectos adversos, proporciones de los dispuestos a que se les realice la prueba
Si se realiza la prueba, ¿se mejora la respuesta clínica o se producen efectos adversos?	Ensayos aleatorizados, estudios de cohortes o de casos y testigos en los que la variable predictiva sea la realización de la prueba y las respuestas incluyan morbilidad, mortalidad o costes relacionados bien con la enfermedad o bien con su tratamiento	Cocientes de riesgo, cocientes de posibilidades, cocientes de riesgos instantáneos, número que es necesario tratar, incidencias y cocientes de respuestas deseables e indeseables

* La mayor parte de los estadísticos de esta tabla deben presentarse con intervalos de confianza.

Puede producirse sesgo de espectro por un espectro inadecuado de resultados de la prueba, así como por un espectro inadecuado de la enfermedad. Por ejemplo, considere un estudio de acuerdo interobservador en radiólogos que interpretan mamografías. Si se les pide que clasifiquen las imágenes como normales o anormales, su acuerdo será mucho mayor si las imágenes «positivas» que selecciona el investigador para su estudio se eligieron porque son claramente anormales, y las imágenes «negativas» se seleccionan porque están libres de cualquier anomalía sospechosa.

- Importancia del enmascaramiento. Muchos estudios de pruebas diagnósticas conllevan juicios, como considerar que una radiografía está alterada o si un paciente cumple los criterios para diagnosticar una enfermedad particular. Siempre que sea posible, los investigadores deben ocultar otra información sobre el paciente al que se está estudiando a las personas que interpreten los resultados de la prueba. En un estudio de la contribución de la ecografía al diagnóstico de la apendicitis, por ejemplo, los profesionales que interpreten las ecografías no deben conocer los resultados de la historia clínica ni de la exploración física [1] . Igualmente, el anatomopatólogo que realice la determinación final de quién tiene apendicitis y quién no (el método de referencia con el que se compararán los resultados de la ecografía) no debe saber los resultados de la exploración ecográfica. El enmascaramiento evita que el sesgo, las ideas preconcebidas y la información de otras fuentes distintas a la prueba afecten a estos juicios.

- Fuentes de variación, posibilidad de generalizar y esquema de muestreo. Para algunas preguntas de la investigación, las diferencias entre los pacientes son la principal fuente de variación en los resultados de una prueba. Por ejemplo, algunos lactantes con bacteriemia (bacterias en la sangre) tendrán leucocitosis, mientras que otros no. No se espera que la proporción de lactantes con bacteriemia y con leucocitosis varíe mucho según el laboratorio que realice el hemograma. Por otro lado, los resultados de muchas pruebas dependen de la persona que realiza las pruebas o del entorno en que se han realizado. Por ejemplo, la sensibilidad, la especificidad y la fiabilidad interobservador para la interpretación de las mamografías depende de la capacidad y de la experiencia del lector, así como de la calidad del equipo. Cuando la exactitud puede variar de un lector a otro o de un centro a otro, es útil estudiar a diferentes lectores e instituciones para poder evaluar la homogeneidad de los resultados.

- Método de referencia para el diagnóstico. Algunas enfermedades tienen un método de referencia que generalmente se acepta que indica la presencia o ausencia de la enfermedad en estudio, como el estudio anatomopatológico de una muestra de una biopsia

hística para diagnosticar cáncer. En otras enfermedades, el método de referencia es una definición, como definir la enfermedad coronaria como una obstrucción del 50 % de al menos una arteria coronaria principal observada mediante angiografía coronaria. Otras, como las enfermedades reumatológicas, precisan que un paciente tenga un número mínimo de síntomas, signos o alteraciones analíticas para cumplir los criterios que definen la enfermedad. Por supuesto, si cualquier síntoma, signo o prueba de laboratorio utilizados para diagnosticar una enfermedad se utilizan como parte del método de referencia, un estudio que los compare con ese método de referencia puede hacer que parezcan falsamente buenos. Esto se denomina sesgo de incorporación, porque se *incorpora* la prueba que se está estudiando al método de referencia; evitar este sesgo es uno de los motivos que se han señalado previamente para el enmascaramiento.

También es importante considerar si el método de referencia es realmente de referencia. Si el método de referencia es imperfecto, puede hacer que una prueba parezca peor de lo que realmente es (si en realidad la prueba tiene mejor rendimiento que el método de referencia) o mejor de lo que realmente es (si la prueba índice comete los mismos errores que el método de referencia).

- ¿Qué constituye una prueba positiva? Particularmente, si una prueba tiene resultados continuos (como la concentración sérica de eritropoyetina), un investigador puede sucumbir a la tentación de analizar todos los resultados de las personas con la variable respuesta (p. ej., anemia de las enfermedades crónicas) y de personas que no tienen dicha variable (otros tipos de anemia), y después seleccionar el mejor valor de corte para definir una prueba positiva. Sin embargo, esto es un tipo de sobreajuste (es decir, la variación aleatoria de la muestra particular que se estudia hace que el rendimiento de la prueba parezca mejor de lo que es en la población). Otros abordajes mejores son basar el valor de corte en el conocimiento clínico o biológico de otros estudios, o dividir las pruebas continuas en intervalos y después calcular el cociente de verosimilitudes de cada uno de los intervalos (v. el texto siguiente). Para minimizar el sobreajuste, los valores de corte para definir los intervalos se deben especificar por adelantado, o se deben utilizar números redondos razonables. El sobreajuste es un problema importante a la hora de elaborar reglas de predicción clínica, que se discuten más adelante en este capítulo.

Volver al principio

▪ ESTUDIOS DE REPRODUCIBILIDAD DE LAS PRUEBAS

A veces, los resultados de las pruebas varían según cuándo o dónde se realizaron, o quién las realizó. La variabilidad intraobservador describe la falta de reproducibilidad de los resultados cuando el mismo observador o laboratorio realiza la prueba en la misma muestra en momentos diferentes. Por ejemplo, si un radiólogo observa la misma radiografía en dos ocasiones, ¿qué porcentaje de veces estará de acuerdo consigo mismo en la interpretación, asumiendo que no recuerda su interpretación previa? La variabilidad interobservador describe la falta de reproducibilidad entre dos o más observadores: si otro radiólogo observa la misma radiografía, ¿qué probabilidad hay de que esté de acuerdo con el primer radiólogo?

Muchas veces, el nivel de reproducibilidad (o su ausencia) es la principal pregunta de la investigación. En otros casos, la reproducibilidad se estudia con el fin de mejorar la calidad, ya sea como parte de la asistencia clínica o como parte de un estudio de investigación. Cuando la reproducibilidad es baja (porque la variabilidad intraobservador o interobservador es grande), no es probable que una determinación sea útil, y puede que necesite mejorarse o abandonarse.

Los estudios de reproducibilidad en sí mismos abordan la precisión, no la exactitud ni la validez (cap. 4), por lo que todos los observadores pueden estar de acuerdo entre sí y estar equivocados. Cuando se dispone de un método de referencia, los investigadores de la reproducibilidad intraobservador e interobservador pueden comparar las observaciones realizadas en los sujetos con un método de referencia para determinar la exactitud. Cuando no se dispone de método de referencia, los investigadores deben utilizar los otros métodos para evaluar la validez que se describen en el capítulo 4.

Diseños

El diseño básico para evaluar la reproducibilidad de las pruebas supone comparar los resultados de pruebas de más de un observador o que se han realizado en más de una ocasión. En pruebas que conllevan varias etapas, y en las que diferencias en cualquiera de esas fases podrían afectar a la reproducibilidad, el investigador deberá decidir la amplitud del objetivo del estudio. Por ejemplo, medir el acuerdo interobservador de anatomopatólogos en un conjunto de portaobjetos de frotis de citologías cervicovaginales en un solo hospital puede sobrestimar la reproducibilidad general de los frotis de las citologías cervicovaginales, porque no se detectaría la variabilidad en el modo de obtención de la muestra y en la preparación del frotis.

La magnitud en la que un investigador debe aislar los pasos que podrían producir desacuerdos entre observadores depende, en parte, de los objetivos de su estudio. La mayor parte de los estudios deben calcular la reproducibilidad de todo el proceso de la prueba, porque es lo

que determina si merece la pena realizarla. Por otro lado, un investigador que esté desarrollando o mejorando una prueba puede desear centrarse en los pasos específicos que son problemáticos para mejorar el proceso. En cualquier caso, el investigador debe mencionar el proceso exacto para obtener el resultado de la prueba en el manual de instrucciones (caps. 4 y 17), y describirlo después en la sección de metodología cuando comunique los resultados del estudio.

Análisis

- Variables categóricas. La medida más sencilla del acuerdo interobservador es el porcentaje de observaciones en las que los observadores están exactamente de acuerdo. Sin embargo, cuando las observaciones no están distribuidas homogéneamente en las distintas categorías (p. ej., cuando la proporción de resultados «anormales» en una prueba dicotómica no es próxima al 50 %), puede ser difícil interpretar el acuerdo porcentual, porque no tiene en consideración el acuerdo que podría deberse simplemente a que dos observadores tienen algunos conocimientos sobre la prevalencia de la alteración. Por ejemplo, si el 95 % de los participantes son normales, dos observadores que escogen al azar qué 5 % de las pruebas van a considerar «anómalo» estarán de acuerdo en que los resultados son «normales» alrededor del 90 % de las veces. El acuerdo porcentual también es una medida subóptima cuando una prueba tiene más de dos resultados posibles que están ordenados intrínsecamente (p. ej., normal, limítrofe, anormal), porque el desacuerdo parcial (p. ej., normal/limítrofe) cuenta igual que el desacuerdo completo (normal/anormal).

  Una medida mejor del acuerdo interobservador, denominada kappa (κ) (apéndice 12A), determina la magnitud del acuerdo más allá del que cabría esperar por el conocimiento por los observadores de la prevalencia de la alteración [2] , y puede tener en consideración el acuerdo parcial. El valor de κ oscila entre -1 (desacuerdo perfecto) y 1 (acuerdo perfecto). Un valor de 0 indica que no hay más acuerdo de lo que cabría esperar por las estimaciones de los observadores de la prevalencia de cada nivel de alteración. En general, se considera que valores de κ superiores a 0,8 son muy buenos; se considera que los valores entre 0,6 y 0,8 son buenos.

- Variables continuas. Las medidas de la variabilidad interobservador para variables continuas dependen del diseño del estudio. Algunos estudios miden el acuerdo entre dos máquinas o métodos (p. ej., temperaturas obtenidas con dos termómetros diferentes). La mejor forma de describir los datos de un estudio de este tipo es obtener los datos de los pares de mediciones (cada par está formado por dos mediciones realizadas casi al mismo tiempo en el mismo sujeto) y describir la diferencia media entre sus pares con alguna medida de la dispersión de las variables, como la desviación típica o con qué frecuencia la diferencia supera un umbral clínicamente importante. Por ejemplo, si se cree que una diferencia clínicamente importante de la temperatura corporal es de 0,3 °C, un estudio que compare temperaturas de termómetros timpánicos y rectales podría calcular la diferencia media (± desviación típica) entre las dos técnicas, y describir con qué frecuencia las dos determinaciones difieren en más de 0,3 °C [3] .

  En otros estudios se examina la variabilidad entre métodos de análisis, observadores o instrumentos de estudio en un gran grupo de técnicos, laboratorios o máquinas diferentes. Estos resultados suelen resumirse usando el coeficiente de variación (CV), que es la desviación típica de todos los resultados obtenidos de una sola muestra dividida por el valor medio. Muchas veces se comparan los CV de dos o más métodos o instrumentos diferentes; el que tiene el menor CV es el más preciso (aunque puede no ser el más exacto).

Volver al principio

▪ ESTUDIOS DE LA EXACTITUD DE LAS PRUEBAS

Los estudios de esta sección abordan la pregunta «¿En qué medida la prueba proporciona la respuesta correcta?» Por supuesto, se asume que hay un método de referencia para mostrar cuál es la respuesta correcta.

Diseños

- Muestreo. Los estudios de pruebas diagnósticas pueden tener diseños análogos a los estudios de casos y testigos o los estudios transversales. En el diseño de un estudio diagnóstico de casos y testigos, se muestrea por separado a los que tienen y los que no tienen la enfermedad, y se comparan los resultados de la prueba en los dos grupos.

  Los estudios de pruebas en los que se obtienen muestras por separado de los que tienen y los que no tienen la enfermedad están sujetos a sesgo en la medición o la notificación del resultado de la prueba, porque su medición necesariamente se produce después de la medición del estado de la enfermedad. Además, los estudios con este esquema de muestreo habitualmente no pueden usarse para calcular los valores predictivos (se comenta en el texto siguiente).

  Una muestra consecutiva de pacientes a los que se evalúa para un determinado diagnóstico proporcionará, generalmente,

resultados más válidos e interpretables, incluyendo los valores predictivos. Por ejemplo, Tokuda y cols. (3) observaron que la intensidad de los escalofríos (como sentir frío o sacudidas en todo el cuerpo bajo una manta gruesa) era un buen factor predictivo de bacteriemia en una serie de 526 pacientes adultos y febriles consecutivos en el servicio de urgencias. Como se incluyó a los participantes antes de conocerse si tenían o no bacteriemia, el espectro de pacientes de este estudio debe ser razonablemente representativo de los pacientes que acuden a los servicios de urgencias con fiebre.

Un método de muestreo que denominamos prueba en tándem se usa a veces para comparar entre sí dos pruebas (supuestamente imperfectas). Ambas pruebas se realizan en una muestra representativa de pacientes, y el método de referencia se aplica selectivamente a los pacientes con resultados positivos en una o en ambas pruebas. El método de referencia también debe aplicarse a una muestra aleatoria de pacientes con resultados negativos concordantes para asegurarse de que realmente no tienen la enfermedad. Este diseño, que permite al investigador determinar qué prueba es más exacta sin el gasto de usar el método de referencia en todos los pacientes con resultados negativos en la prueba, se ha utilizado en estudios que comparan diferentes métodos de citología cervical (4).

Los estudios de pruebas pronósticas necesitan diseños de cohortes. En un diseño prospectivo, la prueba se realiza al principio y se sigue a los pacientes para observar quién presenta la respuesta de interés. Un estudio de cohortes retrospectivo se puede utilizar cuando se dispone de una nueva prueba, como la viremia en pacientes infectados por el VIH, si se dispone de una cohorte definida previamente con muestras de sangre almacenadas. La viremia puede determinarse en la sangre almacenada para observar si predice el pronóstico. El diseño de casos y testigos anidado (cap. 8) es particularmente atractivo si la respuesta de interés es rara y la prueba es cara.

- Variable predictiva: resultado de la prueba. Aunque es más fácil pensar en los resultados de una prueba diagnóstica como positivos o negativos, muchas pruebas tienen resultados categóricos, ordinales o continuos. A fin de aprovechar toda la información disponible en la prueba, los investigadores generalmente deben describir los resultados de las pruebas ordinales o continuas, en lugar de dicotomizarlos como «normal o anormal». La mayor parte de las pruebas son más indicativas de enfermedad si son muy anormales que si son levemente anormales, y tienen un intervalo limítrofe en los que no proporcionan mucha información.

- Variable de respuesta: enfermedad (o su respuesta). La variable de respuesta en el estudio de una prueba diagnóstica es la presencia o la ausencia de la enfermedad, que en situación óptima se determina mediante un método de referencia. Si es posible, la evaluación de la respuesta no debe verse influida por los resultados de la prueba diagnóstica que se está estudiando. La forma óptima de realizarlo es ocultando la información a las personas que realizan la prueba de referencia, para que no conozcan los resultados de la prueba índice.

  A veces, particularmente en el caso de pruebas de cribado, la aplicación uniforme del método de referencia no es ética ni viable. Por ejemplo, Smith-Bindman y cols. (5) estudiaron la exactitud de la mamografía según las características del radiólogo que las interpreta. En las mujeres con mamografías positivas se realizaron pruebas adicionales, finalmente con evaluación anatomopatológica como método de referencia. Sin embargo, no es razonable realizar biopsias mamarias a mujeres cuyas mamografías son negativas. Por tanto, para determinar si estas mujeres tienen mamografías falsamente negativas, los autores relacionaron sus resultados de las mamografías con registros de tumores locales, y consideraron que el método de referencia era el hecho de si se diagnosticó o no cáncer de mama en el año que siguió a la mamografía. Esta solución supone que todos los cánceres de mama que existen en el momento de realizar la mamografía serán diagnosticados en 1 año, y que todos los cánceres de mama diagnosticados en 1 año ya estaban presentes en el momento de la mamografía. La medición del método de referencia de forma diferente, dependiendo del resultado de la prueba, genera la posibilidad de sesgo, que se discute con más detalle al final del capítulo, aunque en ocasiones es la única opción posible.

  La variable de respuesta en un estudio de una prueba pronóstica supone lo que les ocurre a los pacientes con una enfermedad, como el tiempo que viven, qué complicaciones presentan o qué tratamientos adicionales precisan. De nuevo, es importante el enmascaramiento, especialmente si los médicos que atienden a los pacientes pueden tomar decisiones basadas en los factores pronósticos que se están estudiando. Por ejemplo, Rocker y cols. (6) observaron que las estimaciones del pronóstico que realizaban los médicos que atendían a los pacientes, pero no las de las enfermeras clínicas, se asociaban independientemente a la mortalidad en la unidad de cuidados intensivos. Podría deberse a que los médicos tenían más experiencia en la estimación de la gravedad de la enfermedad, pero también podría deberse a que las estimaciones del pronóstico de los médicos tenían mayor efecto que las de las enfermeras en las decisiones de retirar las medidas de apoyo vital. Para distinguir entre estas posibilidades, sería útil obtener estimaciones del pronóstico de otros médicos, además de los que intervienen en la toma o la ejecución de las decisiones sobre la retirada de las medidas de apoyo.

Análisis

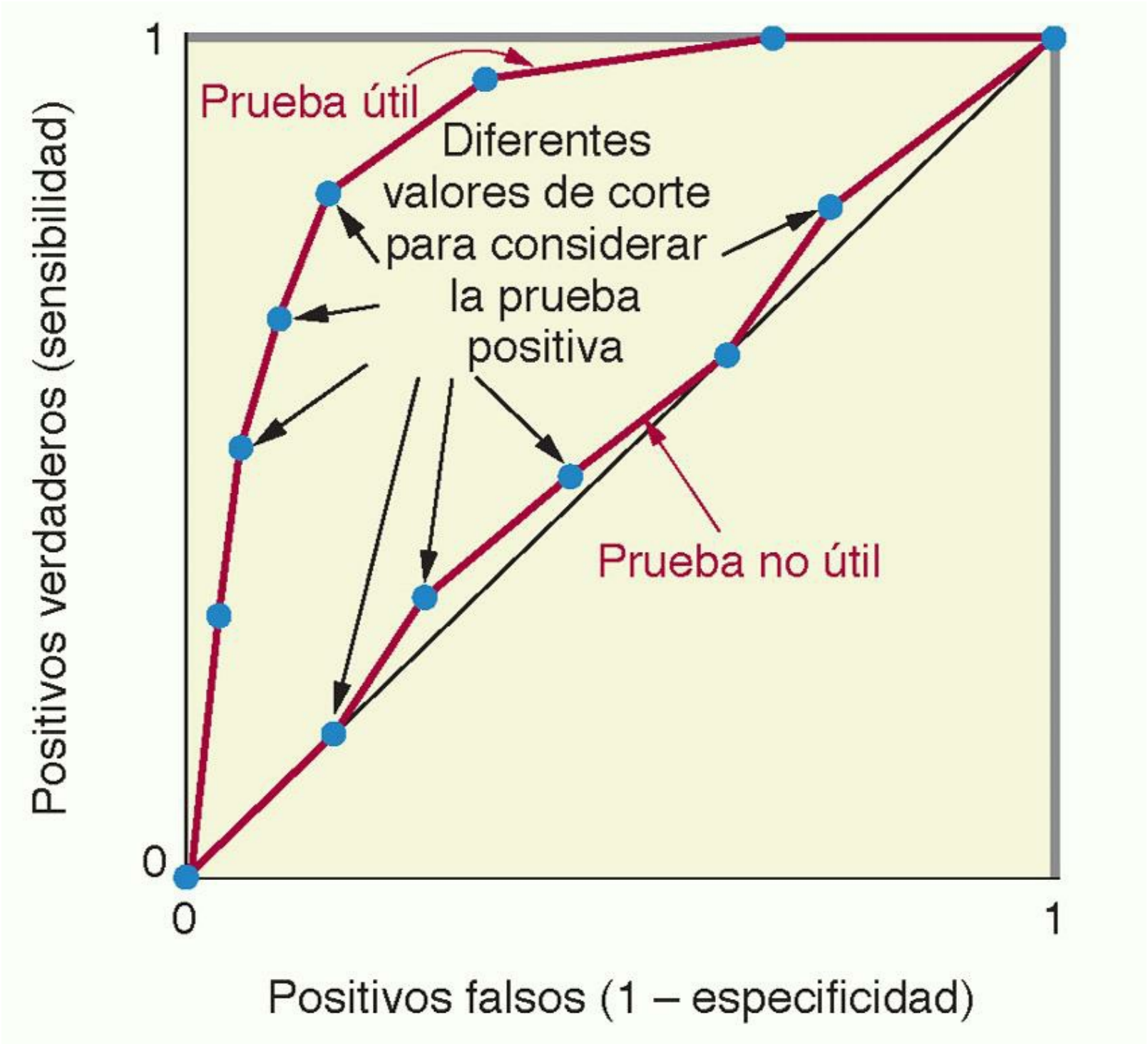
- Sensibilidad, especificidad y valores pronóstico positivos y negativos. Cuando se comparan los resultados de una prueba dicotómica con un método de referencia dicotómico, los resultados pueden resumirse en una tabla de 2 × 2 (tabla 12-2). La sensibilidad de una prueba se define como la proporción de participantes con la enfermedad en los que la prueba da la respuesta correcta (es decir, es positiva), mientras que la especificidad es la proporción de participantes sin la enfermedad en los que la prueba da la respuesta correcta (es decir, es negativa). Si la muestra de pacientes a los que se estudia es representativa del grupo de pacientes en los que se utilizaría la prueba, se pueden calcular dos parámetros adicionales. El valor predictivo positivo es la proporción de personas con un resultado positivo en la prueba que tienen la enfermedad; el valor predictivo negativo es la proporción de personas con un resultado negativo en la prueba que no tienen la enfermedad.

- Curvas de eficacia diagnóstica. Muchas pruebas diagnósticas proporcionan resultados ordinales o continuos. Con estas pruebas puede haber varios valores de sensibilidad y especificidad, dependiendo del valor de corte elegido para definir una prueba positiva. Esta concesión mutua entre sensibilidad y especificidad puede representarse usando una técnica gráfica desarrollada originalmente en electrónica: curvas de eficacia diagnóstica o curvas ROC***(receiver operating characteristic)***. El investigador selecciona varios valores de corte y determina la sensibilidad y la especificidad en cada punto. Representa después la sensibilidad (tasa de positivos verdaderos) en el eje de ordenadas (Y) en función de 1 - especificidad (tasa de positivos falsos) en el eje de abscisas (X). Una prueba ideal es una que alcance la esquina superior izquierda de la gráfica (un 100 % de positivos verdaderos y ningún positivo falso). Una prueba sin valor sigue la diagonal desde la esquina inferior izquierda hasta la superior derecha: en cualquier valor de corte, la tasa de positivos verdaderos es la misma que la de positivos falsos (fig. 12-1). El área bajo la curva ROC, que, por tanto, oscila desde 0,5 para una prueba fútil hasta 1 para una prueba perfecta, es un resumen útil de la exactitud general de una prueba, y puede utilizarse para comparar la exactitud de dos o más pruebas.

- Cocientes de verosimilitudes. Aunque la información en una prueba diagnóstica con resultados continuos u ordinales puede resumirse usando curvas ROC de sensibilidad y especificidad, existe una forma mejor. Los cocientes de verosimilitudes permiten que el investigador se beneficie de toda la información de una prueba. Para cada resultado de una prueba, el cociente de verosimilitudes es la proporción de la posibilidad de ese resultado en alguien con la enfermedad con respecto a la posibilidad de ese resultado en alguien sin la enfermedad.

		MÉTODO DE REFERENCIA			
		ENFERMEDAD	**SIN ENFERMEDAD**	**TOTAL**	
Prueba	*Positiva*	a Positiva verdadera	b Positiva falsa	a + b	Valor predictivo positivo* a/(a + b)
	Negativa	c Negativa falsa	d Negativa verdadera	b + d	Valor predictivo negativo* d/(c + d)
	Total	a + b	b + d		
		Sensibilidad a/(a + c)	Especificidad d/(b + d)		

* Los valores predictivos positivo y negativo pueden calcularse a partir de una tabla de 2 × 2 como esta cuando la prevalencia de la enfermedad es (a + c)/(a + b + c + d). No será así si se muestrean por separado los participantes con y sin enfermedad (p. ej., 100 de cada tipo en un estudio con muestreo de casos y testigos).



▪FIGURA 12-1. Curvas de eficacia diagnóstica (curvas ROC) para pruebas útiles y no útiles.
Herramientas de imágenes

P se lee «probabilidad de» y el símbolo «|» se lee «dado». Así, *P* (resultado|enfermedad) es la probabilidad del resultado dada la enfermedad, y *P* (resultado|ausencia de enfermedad) es la probabilidad de ese resultado dada la ausencia de enfermedad. El cociente de verosimilitudes es una razón de estas dos probabilidades [4] .

Cuanto mayor sea el cociente de verosimilitudes, mejor será la prueba para *confirmar* la enfermedad; un cociente de verosimilitudes superior a 100 es muy alto (y poco habitual en las pruebas). Por otro lado, cuanto menor sea el cociente de verosimilitudes (cuanto más cercano sea a 0), mejor será la prueba para *descartar* la enfermedad. Un cociente de verosimilitudes de 1 (en el intervalo de 0,8 a 1,25) ofrece poca información útil.

Un ejemplo de cómo calcular el cociente de verosimilitudes se muestra en la tabla 12-3 , que presenta resultados de un estudio de hemograma completo en recién nacidos con riesgo de infecciones graves (7). Un recuento leucocitario inferior a 5 000 leucocitos/ml era más frecuente en lactantes con infecciones graves que en otros lactantes. El cálculo del cociente de verosimilitudes sencillamente cuantifica esto: el 19 % de los lactantes con infecciones tenía recuentos leucocitarios menores de 5 000 leucocitos/ml, en comparación con solo el 0,52 % de los que no tenían infecciones. Por lo tanto, el cociente de verosimilitudes es 19 %/0,52 % = 36.

RECUENTO LEUCOCITARIO	INFECCIÓN GRAVE		COCIENTE DE VEROSIMILITUDES
	SÍ	NO	
< 5 000	46 19 %	347 0,52 %	36
5 000-9 999	53 22 %	5 103 7,6 %	2,9
10 000-14 999	53 22 %	16 941 25 %	0,86
15 000-19 999	45 18 %	21 168 31 %	0,58
≥ 20 000	48 20 %	23 818 35 %	0,56
Total	245 100 %	67 377 100 %	

- Riesgo absoluto, cocientes de riesgo, diferencias de riesgo y cocientes de riesgos instantáneos. El análisis de estudios de pruebas pronósticas es similar al de otros estudios de cohortes. Si se realiza un seguimiento durante un período de tiempo establecido (p. ej., 3 años) a todos los que participan en un estudio de una prueba pronóstica, con pocas pérdidas durante el seguimiento, los resultados pueden resumirse con riesgos absolutos, cocientes de riesgo y diferencias de riesgo. Especialmente, cuando el seguimiento es completo y dura poco, los resultados de los estudios de las pruebas pronósticas se resumen a veces como los de las pruebas diagnósticas, usando la sensibilidad, la especificidad, el valor predictivo, los cocientes de verosimilitudes y las curvas ROC. Por otro lado, cuando se sigue a los participantes del estudio durante períodos de tiempo variables, es preferible realizar una técnica de análisis de supervivencia que tenga en consideración la duración del período de seguimiento y que estime los cocientes de riesgos instantáneos (8).

- Mejoría neta por la reclasificación. Para nuevas pruebas o biomarcadores cuya finalidad es la predicción de episodios futuros de la enfermedad, es importante cuantificar lo que añaden las nuevas pruebas a los modelos de predicción existentes. Aunque una forma de hacerlo es analizar la cantidad en la que aumentan el área bajo la curva ROC, los cambios del área bajo la curva ROC a menudo son pequeños, incluso para factores predictivos bien establecidos, y son difíciles de traducir en los cambios previstos de las decisiones clínicas y la evolución de los pacientes (9 , 10). Un abordaje más directo, que tiene su máxima utilidad cuando los umbrales para el tratamiento están bien establecidos, es examinar con qué frecuencia un modelo o una regla de predicción clínica que incluye la nueva prueba modifica la clasificación de los pacientes desde una categoría de riesgo (con su consiguiente decisión de tratamiento) hasta otra, en comparación con el modelo antiguo. Si la nueva prueba mejora la predicción, *ascenderán* hasta una categoría de riesgo superior más pacientes que presentan la variable de respuesta («casos») que pacientes *descenderán* a una categoría de riesgo inferior; sería cierto lo contrario en las personas que no presentaran la variable de respuesta («testigos»): debería *disminuir* el riesgo en más pacientes que en los que *ascendería*. La mejoría neta por la reclasificación (MNR) cuantifica estas diferencias como sigue (11):

  MNR = *P* (ascenso|caso) - *P* (descenso|caso) + *P* (descenso|testigo) - *P* (ascenso|testigo)

donde *P* (ascenso|caso) es la proporción de casos en los que el modelo con el nuevo marcador llevaría a que el paciente ascendiera a una categoría de riesgo superior, y los otros términos se definen en consecuencia. Por ejemplo, Shepherd y cols. (12) encontraron que la adición del volumen fibroglandular calculado en la mamografía (es decir, la cantidad estimada de tejido mamario con riesgo de neoplasia maligna) a un modelo que incluía los factores de riesgo clínicos tradicionales mejoraba la predicción de la aparición posterior de un carcinoma de mama o un carcinoma ductal *in situ* con un MNR del 21 % (p = 0,0 001).

Volver al principio

▪ ESTUDIOS PARA ELABORAR REGLAS DE PREDICCIÓN CLÍNICA

Los estudios para elaborar reglas de predicción clínica difieren de los estudios de pruebas (reglas) ya existentes, porque el objetivo es mejorar las decisiones clínicas utilizando métodos matemáticos para elaborar una nueva prueba (compuesta), en lugar de evaluar una que ya existe.

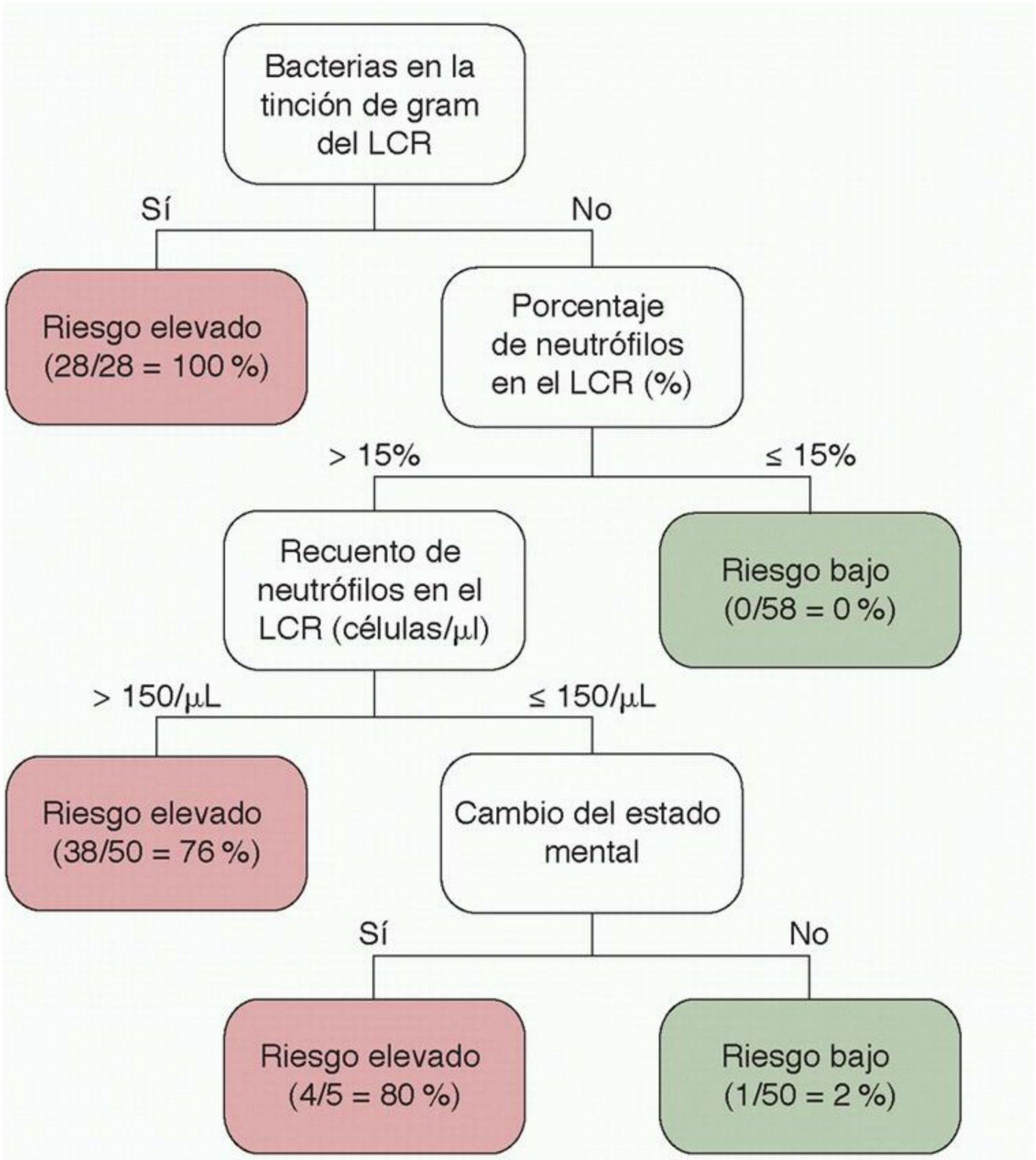
Los participantes en estos estudios deben ser similares a las personas a las que se aplicará la regla. Es probable que las reglas de predicción clínica tengan su máxima utilidad cuando su objetivo es guiar una decisión clínica específica, como la decisión de iniciar el tratamiento con estatinas (para la cual se utiliza la puntuación de riesgo de Framingham). Por lo tanto, los pacientes deben ser aquellos en los que se deben tomar decisiones clínicas específicas, especialmente aquellos en los que actualmente es difícil o hay incertidumbre (13). Muchos estudios para elaborar reglas de decisión clínica incluyen a participantes de un único centro, aunque los que se han elaborado utilizando datos de múltiples centros tienen más probabilidad de poderse generalizar.

Los modelos matemáticos para elaborar reglas de predicción generalmente suponen el uso de una técnica multivariada para seleccionar variables predictivas candidatas y combinar sus valores para generar una predicción. Entre las variables candidatas deben estar todas las variables predictivas conocidas y creíbles que se puedan medir de manera fácil, fiable y económica. Un modelo multivariado, como el análisis de regresión logística o el modelo de Cox (de los riesgos proporcionales), puede cuantificar la contribución independiente de las variables predictivas candidatas a la predicción de la variable de respuesta. Las que tengan una asociación más fuerte y constante con la variable de respuesta se pueden incluir en la regla, y se pueden asignar puntos a diferentes valores de las variables predictivas dependiendo de los coeficientes del modelo. Por ejemplo, Wells y cols. (14) utilizaron análisis de regresión logística en 40 posibles variables predictivas clínicas de embolia pulmonar para elaborar una puntuación predictiva basada en tan solo siete variables (tabla 12-4). Esta puntuación, actualmente popular, se utiliza para asignar una probabilidad previa de embolia pulmonar a fin de guiar las decisiones diagnósticas posteriores y la interpretación de sus resultados (15).

Una técnica alternativa, que no precisa modelado y es útil para generar reglas con sensibilidad elevada, son las particiones recursivas, o análisis de árbol de clasificación y regresión (AACR). Esta técnica genera un árbol que plantea una serie de preguntas que se responden con sí/no, y lleva al usuario hacia abajo por diferentes ramas, dependiendo de las respuestas. Al final de cada rama habrá una probabilidad estimada de la variable de respuesta. Se puede diseñar el árbol para que tenga una sensibilidad elevada, indicando al programa informático que haga que la penalización de los resultados negativos falsos sea mayor que la de los resultados positivos falsos. En la figura 12-2 se muestra un ejemplo de un árbol de este tipo que se utiliza para predecir la meningitis bacteriana en adultos con meningitis (16).

CARACTERÍSTICA CLÍNICA	**PUNTOS**
Embolia pulmonar o trombosis venosa profunda previa	+ 1,5
Frecuencia cardíaca > 100 latidos por minuto	+ 1,5
Cirugía o inmovilización reciente (en los 30 días previos)	+ 1,5
Signos clínicos de trombosis venosa profunda	+ 3
Un diagnóstico alternativo es menos probable que la embolia pulmonar	+ 3
Hemoptisis (expectoración de sangre)	+ 1
Cáncer (tratado en los 6 meses previos)	+ 1
PROBABILIDAD CLÍNICA ESTIMADA DE EMBOLIA PULMONAR (15)	**PUNTUACIÓN TOTAL**

Baja (probabilidad ~1 %-2 %)	0-1
Intermedia (probabilidad ~16 %)	2-6
Elevada (probabilidad ~40 %)	≥ 7



▪FIGURA 12-2. Ejemplo de un árbol de clasificación y regresión para distinguir la meningitis bacteriana de la meningitis vírica en adultos

(16). Los recuadros blancos sirven para dividir a los pacientes en los que tienen riesgo elevado de meningitis bacteriana (recuadros rojos) y los que tienen riesgo bajo (recuadros verdes); los números muestran las proporciones con meningitis bacteriana [5] en las «ramas terminales» roja y verde del árbol.
Herramientas de imágenes

Independientemente del método elegido para elaborar la regla, es importante que se valide en un grupo de pacientes diferente al grupo en el que se obtuvo. Un motivo para esto es evitar el sobreajuste (es decir, aprovechar la tendencia de una única muestra al error aleatorio para aumentar la fuerza predictiva de algunos factores). Los sobreajustes se pueden abordar dividiendo la cohorte en los conjuntos de datos de derivación (habitualmente el 50 %-67 % de la muestra) y de validación, y estudiando con los datos de la cohorte de validación la regla obtenida en la cohorte de derivación. Sin embargo, esto valida las reglas únicamente en una población similar a aquella en la que se obtuvo (es decir, aborda únicamente la validez interna). Para abordar la validez externa es importante determinar en qué medida la regla tiene buen rendimiento en diferentes poblaciones («validación prospectiva») (17).

Volver al principio

▪ ESTUDIOS DEL EFECTO DE LOS RESULTADOS DE LAS PRUEBAS EN LAS DECISIONES CLÍNICAS

Una prueba puede ser exacta, pero si la enfermedad es muy rara, la prueba puede ser positiva con una frecuencia tan baja que apenas merece la pena realizarla. Otras pruebas pueden no afectar a las decisiones clínicas, porque no aportan nueva información aparte de la que ya se conocía (p. ej., por la historia clínica y la exploración física). Los diseños de estudios en esta sección abordan el rendimiento de las pruebas diagnósticas y sus efectos sobre las decisiones clínicas.

Tipos de estudios

- Estudios de rendimiento diagnóstico. Estos estudios permiten responder preguntas como:
    - Cuando se solicita una prueba para una determinada indicación, ¿con qué frecuencia es anormal?
    - ¿Puede un resultado anormal de una prueba predecirse a partir de otra información disponible en el momento de realizar la prueba?
    - ¿En qué grupos de pacientes tiene la prueba la máxima y la mínima utilidad?
    - ¿Qué ocurre con los pacientes con resultados anormales? ¿Los beneficios superan a los perjuicios?

  Los estudios de rendimiento diagnóstico calculan la proporción de pruebas positivas entre pacientes con una indicación particular para la realización de la prueba. Lamentablemente, mostrar que una prueba es positiva a menudo no es suficiente para indicar que la prueba debe realizarse. Sin embargo, un estudio de rendimiento diagnóstico en el que se muestra que una prueba casi siempre es negativa puede ser suficiente para cuestionar su uso para esa indicación.

  Por ejemplo, Siegel y cols. (18) estudiaron el rendimiento de los coprocultivos en pacientes hospitalizados con diarrea. Aunque no se realizaron coprocultivos a todos los pacientes con diarrea, parece razonable suponer que los que la sufren, si es que hay alguno, tienen más probabilidad de tener un cultivo positivo que los que no. En total, solo 40 (2 %) de 1 964 coprocultivos fueron positivos. Además, ninguno de los resultados positivos correspondía a los 997 pacientes que habían estado en el hospital durante más de 3 días. Como es poco probable que un coprocultivo negativo afecte al tratamiento en estos pacientes con una probabilidad baja de diarrea bacteriana, los autores concluyeron que los coprocultivos tienen poca utilidad en pacientes con diarrea que han estado en el hospital más de 3 días.

- Estudios de toma de decisiones clínicas de antes/después. Estos diseños abordan directamente el efecto del resultado de una prueba sobre las decisiones clínicas. El diseño suele conllevar una comparación entre lo que los médicos hacen (o dicen que harían) antes y después de obtener los resultados de una prueba diagnóstica. Por ejemplo, Carrico y cols. (19) estudiaron prospectivamente la utilidad de la ecografía abdominal en 94 niños con dolor abdominal inferior agudo. Pidieron a los médicos que habían solicitado las ecografías que registraran su impresión diagnóstica y cuál sería su tratamiento si no dispusieran de una ecografía. Tras realizar las ecografías y proporcionar a los médicos los resultados, les preguntaron de nuevo. Observaron que la información ecográfica cambió el plan inicial del tratamiento en el 46 % de los pacientes.

  Por supuesto (como se comenta más adelante), la alteración de una decisión clínica no garantiza que un paciente se vaya a beneficiar, y algunas decisiones alteradas realmente podrían ser perjudiciales. Los estudios en los que se demuestran los efectos sobre las decisiones tienen su máxima utilidad cuando la evolución natural de la enfermedad y la eficacia del tratamiento están

claras. En el ejemplo anterior, probablemente habría un beneficio al cambiar la decisión de «alta hospitalaria» a «laparotomía» en niños con apendicitis, o de «laparotomía» a «observación» en niños con dolor abdominal inespecífico.

Volver al principio

▪ ESTUDIOS DE VIABILIDAD, COSTES Y RIESGOS DE LAS PRUEBAS

Otra área importante para la investigación clínica se relaciona con los aspectos prácticos de las pruebas diagnósticas. ¿Qué proporción de pacientes devolverá una tarjeta postal con los resultados de pruebas cutáneas de la tuberculosis? ¿Cuáles son los efectos médicos de las pruebas de cribado falsamente positivas en recién nacidos y los efectos psicológicos sobre los padres? ¿Qué proporción de colonoscopias se complican con perforación colónica?

Aspectos del diseño

Los estudios de viabilidad, costes y riesgos de las pruebas suelen ser descriptivos. El esquema de muestreo es importante, porque las pruebas a menudo varían entre las personas o las instituciones que las realizan y en los pacientes a los que se les realizan.

Una opción sencilla es estudiar a todo aquel al que se realice la prueba, como en un estudio de la tasa de devolución de tarjetas postales tras las pruebas cutáneas de la tuberculosis. Por otro lado, para algunas preguntas, los participantes del estudio solo pueden ser aquellos con resultados positivos o positivos falsos. Por ejemplo, Bodegard y cols. (20) estudiaron familias de lactantes que habían obtenido resultados positivos falsos en una prueba de detección de hipotiroidismo en recién nacidos, y observaron que los temores sobre la salud del niño persistieron al menos 6 meses en el 20 % de las familias.

Pueden producirse efectos adversos no solo por resultados positivos falsos, sino también por la propia prueba. Por ejemplo, Rutter y cols. (21) utilizaron una historia clínica electrónica para hacer un estudio de cohortes retrospectivo de acontecimientos adversos graves (perforación, hemorragia y diverticulitis aguda) en los 30 días siguientes a una colonoscopia en los pacientes de la Group Health Cooperative de Puget Sound.

Análisis

Los resultados de estos estudios suelen poder resumirse con estadísticos descriptivos simples, como medias y desviaciones típicas, medianas, recorridos y distribuciones de frecuencia. Las variables dicotómicas, como la incidencia de efectos adversos, pueden resumirse con proporciones y sus intervalos de confianza (IC) del 95 %. Por ejemplo, en el estudio ya mencionado, Rutter y cols. (21) encontraron perforaciones en 21/43 456 colonoscopias, lo que corresponde al 0,48 por 1 000, con un intervalo de confianza del 95 % desde 0,30 hasta 0,74 por 1 000.

Generalmente no hay líneas claras que dividan las pruebas en las que son y no son viables, o en las que tienen o no un riesgo inaceptable de efectos adversos. Por este motivo es útil, en la etapa de diseño del estudio, especificar los criterios para decidir que la prueba es aceptable. ¿Qué tasa de seguimiento será insuficiente? ¿Qué incidencia de complicaciones será demasiado alta?

Volver al principio

▪ ESTUDIOS DEL EFECTO DE LAS PRUEBAS EN LAS RESPUESTAS

El mejor modo de determinar el valor de una prueba médica es observar si los pacientes a los que se ha realizado presentan una mejor respuesta clínica (p. ej., viven más o con mejor calidad de vida) que aquellos a los que no se les ha realizado. Los ensayos aleatorizados son el diseño ideal para hacer esta determinación, pero los ensayos de pruebas diagnósticas son, a veces, difíciles de realizar. El valor de las pruebas suele, por tanto, calcularse a partir de estudios de observación. La diferencia clave entre los diseños descritos en esta sección y los diseños experimentales y de observación comentados en otras secciones de este libro es que la variable predictiva, en este caso, es *la realización de la prueba*, en lugar de un tratamiento, factor de riesgo o resultado de una prueba.

Diseños

No es probable que la prueba en sí tenga ningún beneficio directo sobre la salud del paciente. Solo cuando el resultado de la prueba conduce a intervenciones preventivas o terapéuticas eficaces es cuando el paciente puede obtener beneficio (22). Por tanto, una advertencia importante sobre los estudios de resultados de pruebas es que la variable predictiva que se está estudiando no es solo una prueba (p. ej., una prueba de sangre oculta en heces), sino también toda la asistencia médica que se realiza posteriormente (p. ej., procedimientos para el seguimiento de los resultados anormales, colonoscopia, etc.).

Es mejor si la variable predictiva de estos estudios es una medida de morbilidad o mortalidad, no simplemente un diagnóstico o la fase de la enfermedad. Por ejemplo, mostrar que los varones en los que se ha realizado una detección selectiva de cáncer de próstata presentan una mayor proporción de cánceres diagnosticados en un estadio precoz no establece por sí solo el valor de esa detección (23 , 24). Muchos de esos cánceres no habrían producido ningún problema si no se hubieran detectado.

La respuesta debe ser lo suficientemente amplia para incluir los posibles efectos adversos de la prueba y el tratamiento, y puede incluir también los efectos psicológicos y médicos de la prueba. Por lo tanto, en un estudio del valor de la detección del antígeno prostático específico para el cáncer de próstata, se debería incluir la impotencia y la incontinencia relacionadas con el tratamiento, además de la morbilidad y la mortalidad relacionadas con el cáncer. Cuando se estudia a muchas personas que se espera que puedan beneficiarse (como suele ocurrir), pueden ser importantes los efectos adversos menos graves en los que no sufren la enfermedad, porque se producirán con mucha más frecuencia. Mientras que los resultados negativos de la prueba pueden aliviar y reconfortar a algunos pacientes (25), en otros, los efectos psicológicos de los resultados positivos falsos, la pérdida de los seguros y los efectos secundarios molestos (pero no mortales) de los fármacos preventivos o la cirugía pueden superar a los infrecuentes beneficios (24).

- Estudios de observación. Los estudios de observación suelen ser más rápidos y fáciles, y menos costosos, que los ensayos clínicos. Sin embargo, tienen importantes inconvenientes, especialmente porque los pacientes a los que se estudia tienden a diferir de los que no son estudiados en aspectos importantes que pueden relacionarse con el riesgo de sufrir una enfermedad o con su pronóstico. Por ejemplo, los que se hacen la prueba podrían tener un riesgo relativamente *bajo* de presentar una respuesta sanitaria adversa, porque las personas que se someten voluntariamente a pruebas médicas y tratamientos tienden a estar más sanas que el promedio, un ejemplo del sesgo del voluntario. Por otro lado, los que se realizan la prueba pueden tener un riesgo relativamente alto, porque es más probable que se hagan pruebas a pacientes cuando hay datos que pueden llevarlos a ellos o a sus médicos a estar preocupados por una enfermedad, un ejemplo de confusión por indicación de la prueba (cap. 9).

  Otro problema habitual en los estudios de observación de pruebas es la falta de normalización y documentación de todas las intervenciones o cambios del tratamiento que siguen a los resultados positivos. Si una prueba no mejora la respuesta en un marco concreto, podría ser porque el seguimiento de los resultados anormales no fue bueno, porque los pacientes no cumplieron las intervenciones planificadas o porque la intervención concreta usada en el estudio no era la ideal.

- Ensayos clínicos. El diseño más riguroso para evaluar el beneficio de una prueba diagnóstica es un ensayo clínico, en el que se asigna aleatoriamente a los participantes a que se les haga o no la prueba. Probablemente, el resultado de la prueba se utilice después para orientar el tratamiento médico. Pueden medirse y compararse diversas mediciones en los dos grupos. Los ensayos aleatorizados reducen al mínimo o eliminan la confusión y el sesgo de selección, y permiten medir todas las respuestas relevantes como mortalidad, morbilidad, coste y satisfacción. La normalización de la prueba y el proceso de intervención permite que otros autores puedan reproducir los resultados.

  Desgraciadamente, los ensayos aleatorizados de pruebas diagnósticas a menudo no son prácticos, especialmente para pruebas diagnósticas que ya se utilizan para la asistencia de los pacientes enfermos. Los ensayos aleatorizados suelen ser más viables e importantes para pruebas que podrían usarse en gran cantidad de personas aparentemente sanas, como nuevas pruebas de cribado.

  Sin embargo, los ensayos aleatorizados pueden plantear cuestiones éticas sobre la no realización de pruebas potencialmente útiles. En lugar de asignar aleatoriamente a los pacientes a realizarles una prueba o no, un método para reducir al mínimo esta cuestión ética es asignar aleatoriamente a algunos pacientes a recibir una intervención que aumente el uso de la prueba, como frecuentes recordatorios con tarjetas postales y ayuda para la programación. El análisis principal debe seguir la norma de «intención de tratar», es decir, se debe comparar todo el grupo al que se aleatorizó a recibir la intervención con el grupo de comparación entero. Sin embargo, esta norma tenderá a producir un sesgo conservador; la eficacia observada de la intervención infravalorará la eficacia real de la prueba, porque algunos participantes del grupo testigo recibirán la prueba, y algunos del grupo de la intervención no. Este problema puede abordarse mediante análisis secundarios que incluyan la incidencia de realización de estudios en los dos grupos y que supongan que toda la diferencia de las variables de respuesta entre los dos grupos se debe a diferentes frecuencias de realización de la prueba. Los beneficios reales de realizar pruebas en los participantes como resultado de la intervención pueden calcularse algebraicamente (8 , 28).

EJEMPLO 12-1. Un buen estudio de observación de una prueba de cribado

Selby y cols. (26) realizaron un estudio de casos y testigos anidado en el *Kaiser Permanente Medical Care Program* para determinar si la sigmoidoscopia de cribado reduce el riesgo de muerte por cáncer de colon. Compararon la prevalencia de sigmoidoscopia previa en pacientes que fallecieron de cáncer de colon con testigos que no fallecieron. Observaron una razón de probabilidades ajustada de 0,41 (IC 95 %, 0,25 a 0,69), lo que sugiere que la sigmoidoscopia causaba una disminución de casi el 60 % de la incidencia de fallecimientos por cáncer de recto y de colon distal.

Un posible problema es que los pacientes a los que se realiza una sigmoidoscopia pueden ser bastante diferentes de los que no se la realizan, y esas diferencias podrían asociarse a una diferencia en la incidencia de fallecimientos esperada por cáncer de colon. Para dirigir esta posible confusión, Selby y cols. examinaron la aparente eficacia de la sigmoidoscopia en la prevención de la muerte por cánceres del colon proximal fuera del alcance de la sigmoidoscopia. Si los pacientes a los que se realizó la sigmoidoscopia tuvieran menos probabilidad de fallecer por cáncer de colon por otras razones, la sigmoidoscopia aparentemente les protegería frente a estos cánceres también. Sin embargo, la sigmoidoscopia carecía de efecto sobre la mortalidad por cáncer de colon proximal (razón de posibilidades ajustada = 0,96; IC 95 %, 0,61 a 1,50), lo que sugiere que la confusión no era la razón de la aparente reducción de la mortalidad por cáncer de colon distal. La especificación de criterios de valoración alternativos (¡por anticipado!) que se espera que *no* se asocien a la variable predictiva de interés (cáncer de colon proximal en este caso), y demostrar después que no lo están, puede reforzar mucho la inferencia causal (27).

Análisis

Los análisis de los estudios del efecto de las pruebas sobre la respuesta son los adecuados al diseño específico utilizado: cocientes de probabilidades para estudios de casos y testigos, y cocientes de riesgos o cocientes de riesgos instantáneos para estudios de cohortes y ensayos clínicos. Una forma conveniente para expresar los resultados es proyectar los resultados del procedimiento de estudio a una cohorte extensa (p. ej., de 100 000), y realizar una lista con el número de pruebas iniciales, pruebas de seguimiento, pacientes tratados, efectos secundarios del tratamiento, costes y muertes en los grupos estudiados y no estudiados.

Volver al principio

▪ ERRORES EN EL DISEÑO Y EL ANÁLISIS DE LOS ESTUDIOS DE PRUEBAS DIAGNÓSTICAS

Al igual que en otros tipos de investigación clínica, los errores en el diseño del análisis de los estudios de pruebas diagnósticas también son frecuentes. En el texto siguiente se presentan algunos de los más frecuentes y graves, con los pasos para evitarlos.

Tamaño de la muestra inadecuado

Si la variable de respuesta de un estudio de una prueba diagnóstica es frecuente, es probable que se pueda obtener una muestra de tamaño adecuado. Cuando la enfermedad o la variable de respuesta son infrecuentes, puede ser necesario un número muy grande de personas. Muchas pruebas de laboratorio, por ejemplo, no son caras, y un rendimiento del 1 % o menos podría justificar su realización, especialmente si pueden diagnosticar una enfermedad grave pero tratable. Por ejemplo, Sheline y Kher (29) revisaron retrospectivamente pruebas analíticas sistemáticas en el momento del ingreso, entre ellas una prueba de VDRL *(Venereal Disease Research Laboratory)* para detectar sífilis, en 252 pacientes psiquiátricos, y observaron que las pruebas analíticas identificaron a un paciente con sífilis no sospechada previamente. Si los síntomas de este paciente psiquiátrico se debieran realmente a la sífilis, sería difícil argumentar que no mereció la pena el gasto de 3 186 dólares en las pruebas de VDRL para llegar al diagnóstico. Pero si la prevalencia real de sífilis no sospechada fuera próxima al 0,4 % observado en este estudio, un estudio de este tamaño de muestra fácilmente podría no haber encontrado casos.

Exclusión incorrecta

Cuando se calculan proporciones, es inadecuado excluir participantes del numerador sin excluir participantes similares del denominador. Por ejemplo, en un estudio de análisis de laboratorio habituales en pacientes del servicio de urgencias con convulsiones de nueva aparición (30), 11 de 136 pacientes (8 %) tenían una alteración de laboratorio corregible (p. ej., hipoglucemia) como causa de la convulsión. En 9 de 11 pacientes, sin embargo, se sospechó la alteración por la historia clínica o la exploración física. Los autores, por lo tanto, comunicaron que solo 2 de 136 pacientes (1,5 %) presentaban alteraciones no sospechadas por la historia clínica o la exploración física. Pero, si se excluye del numerador a todos los pacientes con alteraciones sospechadas, se debería haber excluido a los mismos pacientes también del denominador. El denominador correcto para esta proporción es, por tanto, no todos los 136 pacientes a los que se estudió, sino solo aquellos en los que no se sospechó la existencia de alteraciones analíticas por la historia clínica o la exploración física.

Resultados limítrofes o no interpretables

A veces una prueba no puede dar respuesta alguna, como sucede si el análisis falla, la muestra de la prueba se estropea o el resultado de la prueba se encuadra en una zona que no se sabe si es positiva o negativa. No suele ser válido ignorar estos problemas, pero el modo de manejarlos dependerá de la pregunta específica de la investigación y del diseño del estudio. En los estudios que tratan sobre el coste o la incomodidad de las pruebas, los intentos fallidos de realizar las pruebas son, claramente, resultados importantes.

Se debe considerar que los pacientes con estudios de imagen «no diagnósticos» o con resultados limítrofes en una prueba han tenido ese

resultado específico en la prueba. En efecto, esto puede transformar una prueba dicotómica (positiva, negativa) en una ordinal: positiva, indeterminada y negativa. Pueden trazarse las curvas ROC y calcularse los cocientes de verosimilitudes para los resultados «indeterminados», así como para los positivos y los negativos.

Sesgo de verificación: aplicación selectiva de un solo patrón de referencia

Una estrategia de muestreo habitual para los estudios de pruebas médicas consiste en estudiar (prospectiva o retrospectivamente) pacientes a los que se estudia para detectar una enfermedad y en los que también se aplica el método de referencia para el diagnóstico. Sin embargo, esto causa un problema si los hallazgos que se estudian también se usan para decidir a quién se le realiza el método de referencia. Por ejemplo, considere un estudio de factores predictivos de fracturas en niños que acuden al servicio de urgencias con lesiones en el tobillo, en el que solo se incluyó a niños a los que se hicieron radiografías de sus lesiones de tobillo. Si los que tenían un hallazgo concreto (p. ej., tumefacción del tobillo) tenían más probabilidad de que se les hiciera una radiografía, esto podría afectar a la sensibilidad y la especificidad de la tumefacción del tobillo como prueba para diagnosticar una fractura. Este sesgo, denominado sesgo de verificación, se ilustra numéricamente en el apéndice 12B. El sesgo de verificación se puede evitar usando criterios estrictos para la aplicación del método de referencia que no incluyan la prueba o el hallazgo que se está estudiando. Si esto no es práctico, se puede estimar y corregir el exceso de verificación si el método de referencia se puede aplicar a una muestra aleatoria de las personas en las que la prueba es negativa.

Sesgo de verificación diferencial: diferentes métodos de referencia para los que tienen pruebas positivas y negativas

Otra estrategia es utilizar un método de referencia diferente para aquellos pacientes en los que no está indicado el método de referencia habitual. Por ejemplo, se podría incluir a los pacientes con lesiones de tobillo en los que no se ha realizado una radiografía para contacto telefónico varias semanas después de la lesión, y considerar que no han tenido fracturas si se han recuperado sin problemas. Sin embargo, esto puede producir sesgo de verificación diferencial, también denominado sesgo del doble método de referencia (31). Este sesgo se puede producir siempre que el método de referencia difiera en las personas que tengan un resultado positivo y negativo de la prueba. En el estudio ya mencionado de mamografía (5), el método de referencia en las pacientes con mamografía positiva fue la biopsia, mientras que en las que tenían una mamografía negativa fue el seguimiento para ver si se manifestaba un cáncer en el año siguiente. Tener diferentes métodos de referencia para la enfermedad es un problema si los métodos de referencia no siempre dan los mismos resultados, como ocurriría si un cáncer de mama que se podría detectar mediante biopsia en caso de una mamografía negativa no se manifestara en el seguimiento de 1 año en las pacientes con una mamografía negativa.

Otro ejemplo es un estudio de la ecografía para diagnosticar invaginación en niños pequeños (32). A todos los niños con una ecografía indicativa de invaginación se les realizó el método de referencia, un enema opaco. Por el contrario, a la mayoría de los niños con una ecografía negativa se les observó en el servicio de urgencias y se descartó clínicamente la invaginación. En casos de invaginación que se resolvieron espontáneamente, los dos métodos de referencia darían resultados diferentes: el enema opaco sería positivo, mientras que el período de seguimiento sería negativo. En el apéndice 12C se presenta una ilustración numérica del sesgo de verificación diferencial en este estudio.

El sesgo de verificación diferencial se puede evitar aplicando el mismo método de referencia a todos los participantes. Cuando esto no es posible (como en el estudio de mamografía), los investigadores deben intentar por todos los medios utilizar otros estudios (p. ej., estudios autópsicos en los que se examine la prevalencia de cánceres asintomáticos en pacientes que fallecieron por otras causas en un estudio de pruebas de cribado del cáncer), para evaluar hasta qué punto este sesgo puede amenazar la validez del estudio.

Volver al principio

▪ RESUMEN

- Puede valorarse la utilidad de las pruebas médicas usando diseños que abordan una serie de preguntas cada vez más rigurosas (v. tabla 12-1). En la mayor parte, los diseños de observación habituales proporcionan estadísticos descriptivos de las características de pruebas con sus intervalos de confianza.

- Se debe elegir a los participantes de un estudio de una prueba diagnóstica de entre los pacientes que tienen un espectro de enfermedad y ausencia de enfermedad adecuado para la pregunta de la investigación, y que en la mayoría de los casos reflejen el uso previsto de la prueba en la práctica clínica.

- Si es posible, el investigador debe ocultar a las personas que interpretan los resultados de la prueba y determinan el método de referencia el resto de la información de los pacientes en los que se realice la prueba.

- Medir la reproducibilidad de una prueba, incluyendo la variabilidad intraobservador e interobservador, suele ser un buen primer paso en la evaluación de una prueba.
- Los estudios de la exactitud de las pruebas precisan un método de referencia para determinar si un paciente tiene o no tiene la enfermedad o la respuesta que se está estudiando.
- Los resultados de los estudios de la exactitud de pruebas diagnósticas pueden resumirse usando la sensibilidad, la especificidad, los valores predictivos, las curvas ROC y los cocientes de verosimilitudes. Los estudios del valor de pruebas pronósticas pueden resumirse con cocientes de riesgo, cocientes de riesgos instantáneos o mejoría por la reclasificación.
- Los estudios para elaborar nuevas reglas de predicción clínica están sometidos a los problemas de sobreajuste y ausencia de posibilidad de generalización, por lo que hace falta que las nuevas reglas se validen en muestras de población adicionales.
- El diseño más riguroso para estudiar la utilidad de una prueba diagnóstica es el ensayo clínico, en el que se aleatoriza a los participantes a que se les haga la prueba o no se les haga, y con la mortalidad, la morbilidad, el coste y la calidad de vida entre los criterios de valoración.
- Si los ensayos no son éticos o factibles, pueden ser útiles los estudios de observación de los efectos beneficiosos, los perjuicios y los costes, prestando una atención adecuada a los posibles sesgos y factores de confusión.

Volver al principio

[1]Por otra parte, la exactitud de la anamnesis y la exploración física pueden, por sí solas, compararse con la exactitud de la anamnesis y la exploración física más la ecografía.

[2]A menudo se describe κ como la magnitud del acuerdo que rebasa el que cabe esperar por el azar, aunque la estimación del acuerdo esperado por el azar se basa en la prevalencia de la alteración asignada por cada uno de los observadores.

[3]Aunque se utiliza con frecuencia, es mejor evitar el coeficiente de correlación en estudios de la fiabilidad de las pruebas de laboratorio, porque depende mucho de los valores extremos y no permite que los lectores determinen la frecuencia con la que tienen importancia clínica las diferencias entre las dos medidas. También se deben evitar los intervalos de confianza de la diferencia media, porque su dependencia del tamaño de la muestra hace que puedan llevar a error. Un intervalo de confianza estrecho para la diferencia media entre las dos medidas no implica que, en general, haya un acuerdo elevado, únicamente que la diferencia media entre ellas se mide de forma precisa. Se puede ver un análisis adicional de estos aspectos en Bland y Altman (1), y en Newman y Kohn (2).

[4]Para las pruebas dicotómicas, el cociente de verosimilitudes de una prueba *positiva* es

$$\frac{\text{Sensibilidad}}{1 - \text{Especificidad}}$$

y el cociente de verosimilitudes de una prueba *negativa* es

$$\frac{1 - \text{Sensibilidad}}{\text{Especificidad}}$$

En Newman y Kohn (2) se presentan comentarios detallados sobre cómo utilizar los cocientes de verosimilitudes y la información previa (la probabilidad previa de la enfermedad) para estimar la probabilidad de que un paciente tenga la enfermedad después de conocer el resultado de la prueba (la probabilidad posterior). La fórmula es: Posibilidad previa × cociente de verosimilitudes = posibilidad posterior donde las posibilidades previa y posterior se relacionan con sus respectivas

$$\text{Posibilidad} = \frac{P}{1 - P}$$

probabilidades mediante la fórmula

[5]Los números de las figuras incluyen los conjuntos de datos de derivación y validación.

APÉNDICE 12ACálculo de κ para medir el acuerdo entre observadores
NA

Considere a dos observadores que auscultan para detectar un galope S4 en la exploración cardíaca (tabla 12A-1). Lo registran como presente o ausente. La medida más sencilla del acuerdo entre observadores es la proporción de observaciones en las que los dos observadores están de acuerdo. Esta proporción se puede calcular sumando los números a lo largo de la diagonal que va desde la esquina superior izquierda hasta la inferior derecha, y dividiéndolos por el número total de observaciones. En este ejemplo, de 100 pacientes hubo 10 en los que los dos observadores auscultaron un galope y 75 en los que no lo auscultó ninguno de los dos, lo que da un acuerdo de (10 + 75)/100 = 85 %.

	GALOPE OÍDO POR EL OBSERVADOR 1	**EL OBSERVADOR 1 NO OYE GALOPE**	**TOTAL, OBSERVADOR 2**
Galope oído por el observador 2	10	5	15
El observador 2 no oye galope	10	75	85
Total, observador 1	20	80	100

Cuando las observaciones no están distribuidas uniformemente en las distintas categorías (p. ej., cuando la proporción de resultados «anormales» en una prueba dicotómica es muy diferente del 50 %), o cuando hay más de dos categorías, a veces se utiliza otra medida del acuerdo entre observadores, denominada kappa *(κ)*. κ mide la magnitud del acuerdo más allá de lo que cabría esperar solo por el azar, a la vista de los «valores marginales» observados (es decir, los totales de filas y columnas), y varía desde -1 (desacuerdo perfecto) hasta 1 (acuerdo perfecto). Un valor de κ de 0 indica que la magnitud del acuerdo fue exactamente la que cabía esperar por los totales de filas y columnas. κ se estima como:

La proporción «esperada» en cada celda es simplemente la proporción de la fila de esa celda (es decir, el total de la fila dividido por el tamaño de la muestra) por la proporción de la columna de esa celda (es decir, el total de la columna dividido por el tamaño de la muestra). El acuerdo esperado se obtiene sumando las proporciones esperadas en las celdas a lo largo de la diagonal de la tabla, en la que los observadores están de acuerdo.

Por ejemplo, en la tabla 12A-1 , los observadores parecen haberlo hecho bastante bien: han estado de acuerdo en el 85 % de las veces, pero ¿qué tal lo hicieron en comparación con el acuerdo esperado por los totales marginales? Por el azar solo (a la vista de los valores marginales observados) estarían de acuerdo alrededor del 71 % de las veces (20 % × 15 %) + (80 % × 85 %) = 71 %. Como el acuerdo observado fue del 85 %, κ es (85 % - 71 %)/(100 % - 71 %) = 0,48: respetable, aunque algo menos impresionante que el acuerdo del 85 %.

Cuando hay más de dos categorías de resultados de la prueba, es importante distinguir entre variables ordinales, que están ordenadas intrínsecamente, y variables nominales, que no lo están. Con las variables ordinales, κ tal y como se calculó más arriba no permite capturar toda la información de los datos, porque no asigna un valor parcial a las aproximaciones. Para tener en consideración el acuerdo

parcial, debe usarse el estadístico κ ponderado. (Puede verse una discusión más detallada en Newman y Kohn [29].)

APÉNDICE 12BEjemplo numérico del sesgo de verificación
NA

Considere dos estudios en los que se examina la tumefacción del tobillo como factor predictivo de fracturas en niños con lesiones de tobillo. El primer estudio es una muestra consecutiva de 200 niños. En este estudio se hizo una radiografía a todos los niños con lesiones del tobillo. La sensibilidad y la especificidad de la tumefacción del tobillo son del 80 % y del 75 %, como se muestra en la tabla 12B-1 :

	FRACTURA	**SIN FRACTURA**
Tumefacción	32	40
Sin tumefacción	8	120
Total	40	160

Sensibilidad = 32/40 = 80 % Especificidad = 120/160 = 75 %

En el segundo estudio se usa una muestra seleccionada, en la que solo se hizo una radiografía a la mitad de los niños sin tumefacción del tobillo. Por lo tanto, los números de la fila «Sin tumefacción» se reducirán a la mitad. Esto eleva la sensibilidad aparente desde 32/40 (80 %) hasta 32/36 (89 %), y reduce la especificidad aparente desde 120/160 (75 %) hasta 60/100 (60 %), como se muestra en la tabla 12B-2 :

	FRACTURA	**SIN FRACTURA**
Inflamación	32	40
Sin inflamación	4	60
Total	36	100

Sensibilidad = 32/36 = 89 % Especificidad = 60/100 = 60 %

APÉNDICE 12CEjemplo numérico del sesgo de verificación diferencial
NA

En la tabla 12C-1 se muestran los resultados del estudio de la ecografía para diagnosticar invaginación realizado por Eshed y cols. (32):

	INVAGINACIÓN	**SIN INVAGINACIÓN**
Ecografía +	37	7

Ecografía -	3	104
Total	40	111

Sensibilidad = 37/40 = 93 % Especificidad = 104/111 = 94 %

Los 104 participantes con una ecografía negativa a los que se clasificó como «Sin invaginación» realmente incluían 86 a los que se hizo un seguimiento clínico y no se les realizó un enema opaco. Si alrededor del 10 % de estos participantes (nueve niños) hubieran tenido realmente una invaginación que se resolvió espontáneamente, pero que se habría identificado si se hubiera realizado un enema opaco, y a todos los participantes se les hubiera realizado un enema opaco, esos nueve niños se habrían cambiado de negativos verdaderos a negativos falsos, como se muestra en la tabla 12C-2 :

	INVAGINACIÓN	**SIN INVAGINACIÓN**
Ecografía +	37	7
Ecografía -	3 + 9 = 12	104 - 9 = 95
Total	49	102

Sensibilidad = 37/49 = 76 % Especificidad = 95/102 = 93 %

Se produce un efecto similar, aunque menos pronunciado, si algunos niños con estudios positivos tuvieran una invaginación que se habría resuelto espontáneamente si se hubiera dado la oportunidad (31).

BIBLIOGRAFÍA

1. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307-310. Citado aquí

2. Newman TB, Kohn M. Evidence-based diagnosis. New York: Cambridge University Press, 2009:10-38. Citado aquí

3. Tokuda Y, Miyasato H, Stein GH, et al. The degree of chills for risk of bacteremia in acute febrile illness. Am J Med 2005;118(12):1417. Citado aquí

4. Sawaya GF, Washington AE. Cervical cancer screening: which techniques should be used and why? Clin Obstet Gynecol 1999;42(4):922-938. Citado aquí

5. Smith-Bindman R, Chu P, Miglioretti DL, et al. Physician predictors of mammographic accuracy. J Natl Cancer Inst 2005;97(5):358-367. Citado aquí

6. Rocker G, Cook D, Sjokvist P, et al. Clinician predictions of intensive care unit mortality. Crit Care Med 2004; 32(5):1149-1154. Citado aquí

7. Newman TB, Puopolo KM, Wi S, et al. Interpreting complete blood counts soon after birth in newborns at risk for sepsis. Pediatrics 2010;126(5):903-909. Citado aquí

8. Vittinghoff E, Glidden D, Shiboski S, et al. Regression methods in biostatistics: linear, logistic, survival, and repeated measures models, 2nd ed. New York: Springer, 2012. Citado aquí

9. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification

measures. Ann Intern Med 2009;150(11):795-802. Citado aquí

10. Cook NR. Assessing the incremental role of novel and emerging risk factors. Curr Cardiovasc Risk Rep 2010;4(2): 112-119. Citado aquí

11. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27(2):157-172; discussion 207-212. Citado aquí

12. Shepherd JA, Kerlikowske K, Ma L, et al. Volume of mammographic density and risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2011;20(7):1473-1482. Citado aquí

13. Grady D, Berkowitz SA. Why is a good clinical prediction rule so hard to find? Arch Intern Med 2011;171(19): 1701-1702. Citado aquí

14. Wells PS, Anderson DR, Rodger M, et al. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: increasing the models utility with the SimpliRED D-dimer.Thromb Haemost 2000; 83(3):416-420. Citado aquí

15. Wells PS, Anderson DR, Rodger M, et al. Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and d-dimer. Ann Intern Med 2001;135(2):98-107. Citado aquí

16. Tokuda Y, Koizumi M, Stein GH, et al. Identifying low-risk patients for bacterial meningitis in adult patients with acute meningitis. Intern Med 2009;48(7):537-543. Citado aquí

17. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. JAMA 1997;277(6):488-494. Citado aquí

18. Siegel DL, Edelstein PH, Nachamkin I. Inappropriate testing for diarrheal diseases in the hospital. JAMA 1990; 263(7):979-982. Citado aquí

19. Carrico CW, Fenton LZ, Taylor GA, et al. Impact of sonography on the diagnosis and treatment of acute lower abdominal pain in children and young adults. American Journal of Roentgenology 1999;172(2):513-516. Citado aquí

20. Bodegard G, Fyro K, Larsson A. Psychological reactions in 102 families with a newborn who has a falsely positive screening test for congenital hypothyroidism. Acta Paediatr Scand Suppl 1983;304:1-21. Citado aquí

21. Rutter CM, Johnson E, Miglioretti DL, et al. Adverse events after screening and follow-up colonoscopy. Cancer Causes Control 2012;23(2):289-296. Citado aquí

22. Etzioni DA, Yano EM, Rubenstein LV, et al. Measuring the quality of colorectal cancer screening: the importance of follow-up. Dis Colon Rectum 2006;49(7):1002-1010. Citado aquí

23. Welch HG. Should I be tested for cancer? Maybe not, and here's why. Berkeley, CA: University of California Press, 2004. Citado aquí

24. Welch HG, Schwartz LM, Woloshin S. Overdiagnosed: making people sick in pursuit of health. Boston, MA: Beacon Press, 2011. Citado aquí

25. Detsky AS. A piece of my mind. Underestimating the value of reassurance. JAMA 2012;307(10):1035-1036. Citado aquí

26. Selby JV, Friedman GD, Quesenberry CJ, et al. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer [see comments]. N Engl J Med 1992;326(10):653-657. Citado aquí

27. Prasad V, Jena AB. Prespecified falsification end points: can they validate true observational associations? JAMA 2013;309(3):241-242. Citado aquí

28. Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. Clin Pharmacol Ther 1995; 57(1):6-15. Citado aquí

29. Sheline Y, Kehr C. Cost and utility of routine admission laboratory testing for psychiatric inpatients. Gen Hosp Psychiatry 1990;12(5):329-334. Citado aquí

30. Turnbull TL, Vanden Hoek TL, Howes DS, et al. Utility of laboratory studies in the emergency department patient with a new-onset seizure. Ann Emerg Med 1990;19(4):373-377. Citado aquí

31. Newman TB, Kohn MA. Evidence-based diagnosis. New York: Cambridge University Press, 2009:101-102. Citado aquí

32. Eshed I, Gorenstein A, Serour F, et al. Intussusception in children: can we rely on screening sonography performed by junior residents? Pediatr Radiol 2004;34(2):134-137. Citado aquí

CAPÍTULO 13 Investigación utilizando datos existentes

Deborah G. Grady
Steven R. Cummings
Stephen B. Hulley

Muchas preguntas de la investigación pueden responderse rápida y eficazmente usando datos o muestras que ya se han recopilado. Hay tres métodos generales para usar recursos existentes. El análisis de datos secundarios es el uso de datos existentes para investigar preguntas distintas de las principales para las que se reunieron originalmente los datos. Los estudios auxiliares añaden una o más mediciones a un estudio, a menudo en un subgrupo de participantes, para responder a otra pregunta de la investigación distinta. Las revisiones sistemáticas combinan los resultados de múltiples estudios previos de una determinada pregunta de la investigación, a menudo incluyendo el cálculo de un estimador resumen del efecto que tiene mayor precisión que las estimaciones individuales del estudio. Hacer un uso creativo de datos y muestras ya existentes es una forma rápida y eficaz de que los investigadores nuevos con pocos recursos comiencen a responder preguntas de la investigación importantes, obtengan una experiencia valiosa en un campo de investigación y, en ocasiones, tengan un hallazgo publicable en un marco temporal corto.

▪ VENTAJAS E INCONVENIENTES

Las principales ventajas de los estudios en los que se usan datos existentes son la rapidez y la economía. Una pregunta de la investigación que de otro modo podría necesitar mucho tiempo y dinero para investigar puede, a veces, responderse rápidamente y de forma barata. Por ejemplo, en la base de datos del *Study of Osteoporotic Fractures,* un estudio de cohortes prospectivo diseñado originalmente para estudiar los factores de riesgo de fracturas, Yaffe y cols. utilizaron medidas repetidas que se habían hecho de la actividad física y de la función cognitiva para descubrir que las mujeres que caminaban más tenían un riesgo de deterioro cognitivo un 36 % menor que las mujeres que caminaban menos (1).

Los estudios en los que se utilizan datos o muestras ya existentes también tienen desventajas. La selección de la población a estudiar, qué datos se deben recoger, la calidad de los datos obtenidos, y cómo se registraron y midieron las variables están predeterminados. Los datos existentes pueden haberse recogido de una población que no era la ideal (p. ej., solo hombres, en lugar de hombres y mujeres), el método de medición puede no ser el que el investigador hubiera preferido (antecedentes de hipertensión, una variable histórica dicotómica, en lugar de la presión arterial real) y la calidad de los datos puede ser mala (valores ausentes o incorrectos frecuentes). Pueden no haberse medido o registrado importantes variables de confusión y de respuesta. Todos estos factores contribuyen al principal inconveniente de usar datos existentes: el investigador tiene un control escaso o nulo sobre los datos que se han recopilado y el modo de hacerlo.

Volver al principio

▪ ANÁLISIS DE DATOS SECUNDARIOS

Los datos secundarios pueden proceder de historias clínicas, archivos de facturas de asistencia sanitaria, certificados de defunción, bases de datos públicas y otras muchas fuentes, pero otros estudios de investigación, realizados en el centro del investigador o en otros centros, son una de las principales fuentes de datos secundarios. En muchos estudios se recogen más datos de los que los investigadores analizan, y se pueden utilizar estos datos para documentar resultados interesantes que han pasado desapercibidos. El acceso a esos datos está controlado generalmente por el investigadorprincipal del estudio o por un comité directivo; el nuevo investigador deberá, por tanto, buscar información sobre estudios realizados por otros investigadores en los que se puedan haber hecho mediciones importantes para la pregunta de la investigación. Una de las mejores formas en las que un buen mentor puede ser útil al nuevo investigador es informándole y facilitándole el acceso a conjuntos de datos importantes. Actualmente es necesario que la mayor parte de los estudios financiados por los NIH pongan sus datos a disposición del público.

Otras fuentes fructíferas de datos secundarios son grandes series de datos nacionales y regionales a disposición del público y que no tienen un investigador principal. Las bases de datos informatizadas de este tipo son tan variadas como los motivos que pueda tener la gente para recopilar información. Ofreceremos algunos ejemplos que merecen especial mención, y los lectores pueden localizar otros en sus propias áreas de interés.

- Los registros de tumores son agencias financiadas por el gobierno que recogen estadísticas completas sobre la incidencia, el tratamiento y la evolución del cáncer en determinadas áreas geográficas. Estos registros incluyen actualmente a la cuarta parte de la población de Estados Unidos, y se espera que el área que abarcan aumente en los próximos años. Un objetivo de estos registros es proporcionar datos a investigadores externos. Se dispone de datos combinados de los registros en el programa *Surveillance, Epidemiology, and End Results* (SEER). Por ejemplo, se utilizó el registro SEER de diagnósticos de cáncer de mama para encontrar que la incidencia anual de cáncer de mama que expresaba receptores estrogénicos disminuyó un 13 % en mujeres

posmenopáusicas entre 2001 y 2003; esta tendencia fue paralela a la reducción del uso de terapia hormonal en mujeres posmenopáusicas, lo que indica que la interrupción del tratamiento hormonal redujo el riesgo de cáncer de mama (2).

- Los certificados de defunción pueden usarse para seguir la mortalidad de cualquier cohorte. El National Death Index incluye todos los fallecimientos en Estados Unidos desde 1978. Puede usarse para averiguar la situación vital de los participantes en un estudio anterior, o de los que forman parte de otra serie de datos que incluyen importantes variables predictivas. Un ejemplo clásico es el seguimiento de hombres con cardiopatía isquémica a los que se asignó aleatoriamente a dosis elevadas de ácido nicotínico o un placebo para reducir el colesterol sérico en el estudio *Coronary Drug Project*. En ningún estudio se había demostrado nunca el efecto del tratamiento de los lípidos sobre la mortalidad, y no hubo diferencia en la incidencia de fallecimientos al final de los 5 años de tratamiento aleatorizado, pero en un seguimiento de la mortalidad 9 años después usando el *National Death Index* se mostró una mejoría significativa (3). Si una persona vive o no es una información pública, por lo que se dispuso del seguimiento incluso de los hombres que abandonaron el estudio.

  El *National Death Index* puede usarse cuando se conoce el número de la seguridad social o el nombre y la fecha de nacimiento. La comprobación del fallecimiento es completa en el 99 % con este sistema, y puede obtenerse información adicional de los certificados de defunción (fundamentalmente, la causa de la muerte) de los registros estatales. A nivel local y estatal, muchas jurisdicciones cuentan con sistemas estadísticos informatizados, en los que se introducen datos de las personas (como información de certificados de nacimiento y fallecimiento) en el momento en que se reciben.

- El estudio NHANES (*National Health and Nutrition Examination Survey)* es una serie de estudios en los que se evaluó la salud y el estado nutricional de adultos y niños de Estados Unidos. En estos estudios se utiliza la selección aleatoria por conglomerados de base poblacional para identificar una muestra representativa de todo el país, y se incluyen datos autorreferidos por los pacientes (p. ej., datos demográficos, socioeconómicos, dietéticos y conductas relacionadas con la salud), exploración física, pruebas de laboratorio y otras mediciones. Los datos del estudio NHANES pueden ofrecer estimaciones poblacionales de la prevalencia de enfermedades, factores de riesgo y otras variables. Por ejemplo, se midió la densidad mineral ósea (DMO) de la cadera en dos evaluaciones: 1988-1994 y 2005-2006. Se obtuvieron valores para mujeres y hombres de diferentes razas de Estados Unidos, que se utilizaron para definir la «osteoporosis» como 2,5 desviaciones típicas por debajo del valor medio de la DMO de los adultos jóvenes del estudio NHANES (4). Los investigadores también utilizaron las mediciones repetidas para descubrir que la DMO ha mejorado, y que la prevalencia de osteoporosis ha disminuido (5).

Los datos secundarios pueden ser especialmente útiles para estudios de evaluación de patrones de utilización y respuestas clínicas a los tratamientos médicos. Este método puede complementar la información disponible de ensayos aleatorizados y examinar preguntas que los ensayos no pueden responder. Estos tipos de datos existentes incluyen bases de datos administrativas y clínicas electrónicas, como las desarrolladas por Medicare, el Department of Veterans Affairs, el Kaiser Permanente Medical Group y el Duke Cardiovascular Disease Databank, y registros, como el *San Francisco Mammography Registry* y el *National Registry of Myocardial Infarction*. La información de estas fuentes (gran parte de ellas disponible en la Web) puede ser muy útil para estudiar efectos adversos infrecuentes y para evaluar la utilización y la eficacia en la práctica clínica real de una intervención que se ha demostrado que funciona en el marco de un ensayo clínico. Por ejemplo, se usó el estudio *National Registry of Myocardial Infarction* para examinar los factores de riesgo de hemorragia intracraneal tras el tratamiento del infarto agudo de miocardio (IM) con activador del plasminógeno hístico (tPA) recombinante. En el estudio se incluyó a 71 073 pacientes que recibieron tPA; 673 de ellos presentaron hemorragia intracraneal confirmada mediante tomografía computarizada o resonancia magnética. Un análisis multivariado demostró que una dosis de tPA superior a 1,5 mg/kg se asociaba significativamente a la aparición de hemorragia intracraneal en comparación con dosis inferiores (6). Dado que el riesgo total de presentar una hemorragia intracraneal fue inferior al 1 %, un ensayo clínico en el que se recogieran datos primarios para examinar esta respuesta habría sido prohibitivamente grande y caro.

Otra contribución valiosa de este tipo de análisis de datos secundarios es una mejor comprensión de la diferencia entre eficacia y efectividad. El ensayo clínico aleatorizado es el método de referencia para determinar la eficacia de un tratamiento en una población seleccionada en circunstancias muy controladas en entornos clínicos limitados. En la práctica clínica real, no obstante, los pacientes a los que se trata, la elección de los fármacos y la dosis por el médico que trata a los pacientes, y el cumplimiento con la medicación por parte de los pacientes, son mucho más variables. Estos factores pueden hacer que la aplicación del tratamiento a la población general sea menos eficaz que lo que se observa en los ensayos clínicos. La efectividad de los tratamientos en la práctica real puede estudiarse con datos secundarios. Por ejemplo, se ha demostrado que la angioplastia primaria es superior al tratamiento trombolítico en ensayos clínicos de pacientes con IM agudo (7). Pero esto puede ser cierto solo cuando las tasas de éxito de la angioplastia son tan buenas como las que se alcanzan en el contexto del ensayo clínico. En los análisis secundarios de series de datos de la comunidad no se han observado los beneficios de la angioplastia primaria respecto al tratamiento trombolítico (8 , 9).

El análisis de datos secundarios es, a menudo, el mejor método para describir cómo se utilizan los tratamientos en la práctica clínica. Aunque en los ensayos clínicos se puede demostrar la eficacia de un nuevo tratamiento, este beneficio solo puede producirse si los médicos adoptan el tratamiento. Conocer las frecuencias de utilización, abordar la variación regional y el uso en poblaciones específicas

(como los ancianos, minorías étnicas, personas desfavorecidas económicamente y mujeres) puede tener importantes implicaciones en salud pública. Por ejemplo, utilizando datos disponibles públicamente de una muestra aleatoria del 5 % de los beneficiarios de Medicare, se encontró una importante variación regional de la prevalencia de glaucoma diagnosticado después de ajustar posibles factores de confusión, lo que indica un diagnóstico excesivo o insuficiente en algunas regiones del país (10).

Dos o más series de datos existentes pueden también combinarse para responder a una pregunta de la investigación. Los investigadores que estaban interesados en cómo afecta a la salud el servicio militar usaron el sorteo de reclutamiento de 1970 a 1972, en el que se incluyó a 5,2 millones de hombres de 20 años de edad que fueron declarados idóneos para el servicio militar aleatoriamente por fecha de nacimiento (el primer grupo de datos) y la mortalidad posterior basada en los registros de certificados de defunción (la segunda fuente de datos). La variable predictiva (fecha de nacimiento) era una variable indirecta asignada aleatoriamente para el servicio militar durante la época de la guerra de Vietnam. Los hombres a los que se asignó aleatoriamente para que fueran aptos para el reclutamiento presentaron una mortalidad significativamente mayor por suicidio y accidentes de tráfico en los 10 años siguientes (11). El estudio se realizó con muy poco coste, pese a lo cual fue un método menos sesgado para examinar el efecto del servicio militar sobre causas específicas de fallecimiento posterior que otros estudios de este tema con presupuestos mucho mayores.

Cuando no se dispone de datos individuales, pueden ser útiles a veces las series de datos conjuntos. Los datos conjuntos incluyen información únicamente de grupos de personas (p. ej., tasas de mortalidad por cáncer cervical en cada uno de los 50 estados), no de individuos. Con estos datos solo pueden medirse asociaciones entre esos grupos comparando la información del grupo sobre un factor de riesgo (como las ventas de tabaco por región) con la incidencia de una respuesta (cáncer de pulmón por región). Los estudios para detectar asociaciones basados en datos conjuntos se denominan estudios ecológicos.

La ventaja de los datos conjuntos es su disponibilidad. Su principal inconveniente es que las asociaciones son especialmente susceptibles a la confusión: los grupos tienden a diferir entre sí de muchas formas, no solo en relación con la variable predictiva de interés. En consecuencia, las asociaciones observadas en el conjunto no se mantienen necesariamente para el individuo. Por ejemplo, las ventas de cigarrillos pueden ser superiores en estados con incidencias de suicidio elevadas, pero puede que las personas que se suicidan no sean las que más fuman. Esta situación se denomina falacia ecológica. El uso más adecuado de los datos conjuntos es para comprobar la credibilidad de una nueva hipótesis o para generar nuevas hipótesis. Los resultados interesantes pueden buscarse en otro estudio que utilice datos individuales.

Empezar

Tras elegir un tema de investigación y familiarizarse con la bibliografía de ese campo (incluyendo una exhaustiva búsqueda bibliográfica y el asesoramiento de un mentor experto), el paso siguiente será investigar si la pregunta del estudio puede abordarse con un conjunto de datos existente. La ayuda de un compañero experto puede tener enorme valor para buscar una serie de datos apropiada. Un investigador con experiencia tiene áreas de interés definidas en las que está actualizado, y conoce series de datos importantes y a los investigadores que controlan esos datos, tanto en su propio centro como en otros. Esta persona puede ayudar a identificar y acceder a datos adecuados. A menudo, la pregunta del estudio necesita algún ligero cambio (p. ej., modificando la definición de las variables predictiva y de respuesta) para encajar en los datos disponibles.

La mejor solución puede estar muy próxima, una base de datos en el propio centro. Por ejemplo, un docente de la Universidad de California, San Francisco (UCSF), que estaba interesado en la importancia de las lipoproteínas en la cardiopatía isquémica, comunicó que una de las pocas intervenciones conocidas para reducir la concentración de la lipoproteína(a) eran los estrógenos. Sabiendo que se estaba realizando en la UCSF el *Heart and Estrogen/Progestin Replacement Study* (HERS), un importante ensayo clínico de tratamiento hormonal para evitar la enfermedad coronaria, el docente explicó su interés a los investigadores. Como nadie más había pensado específicamente en estudiar la relación entre esta lipoproteína, el tratamiento hormonal y los episodios de cardiopatía isquémica, diseñó un análisis y un plan de publicación. Tras recibir la autorización del director del estudio HERS, trabajó con estadísticos, epidemiólogos y programadores del centro coordinador para realizar un análisis que publicó posteriormente en una revista importante (12).

A veces, una pregunta de estudio puede abordarse de un modo que poco tiene que ver con el estudio original. Por ejemplo, otro compañero de la UCSF estaba interesado en la utilidad del estudio repetido de la citología cervicovaginal de cribado en mujeres de más de 65 años de edad. Se dio cuenta de que la media de edad de las participantes en el ensayo HERS era de 67 años, de que se exigió que las participantes tuvieran una citología cervicovaginal normal para entrar en el estudio, y de que después se realizaran a las participantes frotis cervicovaginales de cribado anualmente durante el seguimiento. Realizando el seguimiento de los resultados de los frotis cervicovaginales, pudo documentar que hubo 110 citologías cervicovaginales anormales de entre las 2 763 mujeres estudiadas durante un período de 2 años, y solo se observó finalmente una mujer con una histología anormal en el seguimiento. Por lo tanto, todos los resultados anormales de la citología cervicovaginal menos uno fueron positivos falsos (13). Este estudio influyó enormemente en las siguientes recomendaciones del grupo de trabajo US Preventive Services Task Force de que no deben realizarse citologías cervicovaginales en mujeres de riesgo escaso de más de 65 años de edad con pruebas anteriores normales.

A veces es necesario aventurarse más lejos. Partiendo de una lista de variables predictivas y de respuesta cuya relación podría ayudar a responder la pregunta de investigación, un investigador puede intentar localizar bases de datos que incluyan esas variables. Algunos estudios tienen páginas web que permiten el acceso gratuito a los datos del estudio sin solicitar permiso. Cuando los datos no están disponibles en Internet, las llamadas telefónicas o los mensajes de correo electrónico a los autores de estudios previos o a funcionarios gubernamentales podrían permitir el acceso a archivos que contienen datos útiles. Es esencial superar la ansiedad por contactar con extraños para pedir ayuda. La mayor parte de las personas son sorprendentemente colaboradoras, ya sea proporcionando datos ellas mismas o indicando otros lugares donde poder intentarlo.

Una vez localizados los datos para responder a la pregunta del estudio, el siguiente reto es lograr la autorización para usarlos. Es una buena idea usar títulos oficiales y el nombre del dominio institucional en la correspondencia o los correos electrónicos, y señalar al mentor como alguien a quien se reconocerá como experto en esa área. Los investigadores jóvenes deben determinar si sus mentores conocen a los investigadores que controlan la base de datos, ya que una presentación puede ser más eficaz que un contacto frío. Generalmente, es más eficaz trabajar con un investigador o un miembro del personal del estudio que esté interesado en el tema de la investigación e implicado en el estudio que tenga los datos de interés. Este investigador puede facilitar el acceso a los datos y garantizar que conoce los métodos de estudio y el modo en que se midieron las variables, y a menudo se convierte en un valioso compañero y colaborador. Los datos de estudios y ensayos clínicos multicéntricos tienen, generalmente, procedimientos claros para conseguir el acceso a los datos, e incluyen la necesidad de una propuesta por escrito, que debe ser autorizada por un comité de análisis o de publicaciones.

El investigador debe especificar claramente la información que busca y confirmar la solicitud por escrito. Muchos estudios tienen directrices para la solicitud de datos que especifican qué datos se solicitan, cómo se realizarán los análisis y los marcos temporales para la realización del trabajo. Es una buena idea que el tamaño de la solicitud sea mínimo, y ofrecerse a pagar el coste de la preparación de los datos. Si el conjunto de datos está controlado por otro grupo de investigadores, el investigador puede sugerir una relación de colaboración. Además de proporcionar un incentivo por compartir los datos, esto puede hacer participar a un coinvestigador que está familiarizado con la base de datos. Es prudente definir pronto y claramente esta relación, incluso quién será el primer autor de las publicaciones planeadas.

Volver al principio

▪ ESTUDIOS AUXILIARES

La investigación con datos secundarios se beneficia de que la mayoría de los datos necesarios para responder a la pregunta del estudio ya están disponibles. En un estudio auxiliar, el investigador añade una o varias determinaciones a un estudio existente para responder a una pregunta diferente. Por ejemplo, en el ensayo HERS del efecto de la hormonoterapia sobre el riesgo de sufrir episodios coronarios en 2 763 ancianas, un investigador añadió la determinación de la frecuencia y la gravedad de la incontinencia urinaria. La adición de un cuestionario breve en la siguiente exploración planificada permitió hacer un extenso ensayo del efecto de la hormonoterapia sobre la incontinencia urinaria, con un tiempo y unos gastos adicionales escasos (14).

Los estudios auxiliares tienen muchas de las ventajas del análisis de datos secundarios con menos limitaciones. Son baratos y eficientes, y el investigador puede diseñar algunas determinaciones auxiliares clave para responder específicamente a la pregunta del estudio. Los estudios auxiliares pueden añadirse a cualquier tipo de estudio, entre ellos los estudios de casos y testigos y los estudios transversales, pero los estudios prospectivos de cohortes extensos y los ensayos clínicos aleatorizados son los más adecuados.

Los estudios auxiliares tienen el problema de que las mediciones pueden aportar más información cuando se añaden antes de iniciar el estudio, y puede ser más difícil para alguien de fuera identificar los estudios en la fase de planificación. Incluso aunque no se midiera una variable al principio, no obstante, una sola determinación durante el estudio o al final de él puede aportar información útil. Añadiendo medidas de función cognitiva al final del ensayo HERS, los investigadores pudieron comparar la función cognitiva de las ancianas tratadas con hormonoterapia durante 4 años con la función cognitiva de las que recibieron el placebo (15).

Una buena oportunidad para los estudios auxiliares la proporcionan los bancos de suero almacenado, ADN, imágenes, etc., que se encuentran en la mayor parte de los grandes ensayos clínicos y estudios de cohortes. La oportunidad de proponer nuevas mediciones usando estas muestras puede ser un método enormemente rentable para responder a una nueva pregunta de la investigación, especialmente si es posible realizar estas determinaciones en un subgrupo de muestras usando un diseño de casos y cohortes o de casos y testigos anidado (cap. 8). En el estudio HERS, por ejemplo, un estudio de casos y testigos anidado en el que se realizaron análisis genéticos en muestras almacenadas, se demostró que el excesivo número de episodios tromboembólicos en el grupo tratado con hormonas no se debía a una interacción con el factor V Leiden (16).

Cómo empezar

Deben buscarse activamente las oportunidades para realizar estudios auxiliares, especialmente los nuevos investigadores que tienen pocos recursos y tiempo. Un buen punto de partida es identificar estudios con preguntas de la investigación que incluyan la variable predictiva o la variable de respuesta de interés. Por ejemplo, un investigador interesado en el efecto de la pérdida de peso sobre el dolor asociado a la artrosis de rodilla podría empezar por identificar estudios que incluyan una buena medición de la artrosis dolorosa (con cuestionarios validados) o bases de datos con registros de sustituciones articulares que también tengan mediciones previas del peso. Además, el investigador puede buscar estudios de intervenciones (como dieta, ejercicio, cambio de conducta o fármacos) para perder peso. Estos estudios se pueden identificar buscando listas de estudios financiados por el gobierno federal (http://clinicaltrials.gov o http://report.nih.gov), contactando con compañías farmacéuticas que fabrican fármacos para perder peso, y hablando con expertos en la pérdida de peso que estén familiarizados con los estudios en realización. Para crear un estudio auxiliar, el investigador debe, sencillamente, añadir una determinación de síntomas de artrosis en una evaluación de seguimiento de participantes incluidos en estos estudios.

Tras identificar un estudio que proporcione una buena oportunidad para realizar mediciones auxiliares, el siguiente paso será lograr la colaboración de los investigadores del estudio. La mayoría de los investigadores considerarán la adición de breves medidas auxiliares a un estudio establecido si abordan una pregunta importante y no interfieren notablemente con la realización del estudio principal. Pueden ser reacios a añadir mediciones que precisen mucho tiempo de los participantes (p. ej., pruebas de función cognitiva) o que sean cruentas y molestas (colonoscopia) o costosas (tomografía por emisión de positrones).

En general, se necesita la autorización formal del investigador principal o del correspondiente comité del estudio para poder añadir un estudio auxiliar. La mayor parte de los estudios extensos y multicéntricos tienen procedimientos establecidos que precisan una solicitud por escrito. Generalmente, el estudio auxiliar propuesto es revisado por un comité, que puede aprobar, rechazar o revisar dicho estudio. Muchas determinaciones auxiliares precisan financiación, y el investigador de este estudio debe encontrar un modo de pagar el coste. Por supuesto, el coste de un estudio auxiliar es mucho menor que el coste que supondría realizar el mismo estudio independientemente. Los estudios auxiliares también son muy adecuados para algunos tipos de financiación por los NIH, que ofrecen una cantidad moderada para las mediciones y los análisis, aunque aportan un apoyo importante para el desarrollo de la carrera profesional (cap. 19). Algunos grandes estudios pueden tener sus propios mecanismos para financiar estudios auxiliares, especialmente si la pregunta es importante y se considera relevante para la agencia que lo financia.

Los inconvenientes de los estudios auxiliares son pocos. Si en el estudio se van a recoger datos de participantes, se pueden añadir nuevas mediciones, aunque las variables que ya se han medido generalmente no pueden cambiarse. En algunos casos puede que haya problemas prácticos para la obtención de la autorización de los investigadores o del patrocinador para realizar el estudio auxiliar, para formar a quienes realizarán las determinaciones o para obtener un nuevo consentimiento informado de los participantes. Antes del inicio del estudio deben aclararse estos aspectos, además de una clara definición de la autoría de los artículos científicos que se produzcan a partir del estudio auxiliar, y de las normas que regulan su preparación y envío.

Volver al principio

▪ REVISIONES SISTEMÁTICAS

Las revisiones sistemáticas identifican un conjunto de estudios ya realizados que analizan una pregunta de la investigación concreta, y evalúan los resultados de estos estudios para llegar a conclusiones sobre un cuerpo de investigación. A diferencia de otros métodos para revisar bibliografía, en las revisiones sistemáticas se usa un abordaje bien definido para identificar todos los estudios relevantes, obtener las características y los resultados de estudios elegibles y, cuando proceda, calcular un estimador resumen de los resultados generales. Los aspectos estadísticos de una revisión sistemática (cálculo de estimadores resumen del efecto y de su varianza, pruebas estadísticas de heterogeneidad y cálculos estadísticos del sesgo de publicación) se denominan metaanálisis.

Una revisión sistemática puede ser una buena oportunidad para un nuevo investigador. Aunque requiere una sorprendente cantidad de tiempo y esfuerzo, generalmente no precisa una financiación importante ni otros recursos. La realización de una buena revisión sistemática precisa que el investigador se familiarice con la bibliografía sobre la pregunta del estudio. Para los nuevos investigadores, este conocimiento detallado de los estudios publicados tiene un valor incalculable. La publicación de una buena revisión sistemática puede convertir también a un nuevo investigador en «experto» en la pregunta de la investigación. Además, los hallazgos, con mayor potencia por el mayor tamaño de la muestra disponible por los estudios combinados y las peculiaridades de los hallazgos de estudios concretos revelados por comparación con los otros, a menudo representan una contribución científica importante. Los hallazgos de la revisión sistemática pueden ser particularmente útiles para desarrollar directrices prácticas.

En la tabla 13-1 se muestran los elementos de una buena revisión sistemática. Puede encontrarse una buena fuente de información sobre los métodos para realizar revisiones sistemáticas de alta calidad en el Manual Cochrane para realizar revisiones sistemáticas (*Cochrane Handbook for Systematic Reviews*) (http://handbook.cochrane.org). Igual que para otros estudios, los métodos para llevar a cabo cada

uno de estos pasos deben describirse en un protocolo por escrito antes de iniciar la revisión sistemática.

La pregunta de la investigación

Una buena revisión sistemática tiene una pregunta de investigación clara y bien formulada que cumple los criterios FINER (cap. 2). La viabilidad depende notablemente de la existencia de un grupo de estudios sobre esa pregunta. Esta debe describir la enfermedad o afección de interés, la población y el contexto, la intervención y el tratamiento de comparación (para los ensayos), y las respuestas de interés. Por ejemplo:

«En pacientes ingresados en una unidad de cuidados intensivos con síndrome coronario agudo, ¿el tratamiento con ácido acetilsalicílico más heparina intravenosa reduce el riesgo de sufrir infarto de miocardio y muerte durante la hospitalización más que el tratamiento con ácido acetilsalicílico solo?»

Esta pregunta de la investigación llevó a un metaanálisis en el que se encontró que la adición de ácido acetilsalicílico a la heparina mejoraba el pronóstico, y que se publicó en una importante revista médica (17) y tuvo consecuencias importantes sobre los patrones de práctica habitual.

Identificar estudios ya realizados

Las revisiones sistemáticas se basan en una búsqueda exhaustiva y no sesgada de estudios ya realizados. Esta búsqueda debe seguir una estrategia bien definida establecida antes de que se conozcan los resultados de los estudios individuales. El proceso de identificar estudios para su posible inclusión en la revisión, y las fuentes para buscar esos artículos, deben documentarse explícitamente antes del estudio. Las búsquedas no deben limitarse a MEDLINE, que puede no incluir referencias en idiomas distintos del inglés. Dependiendo de la pregunta del estudio, deben incluirse bases de datos electrónicas, como AIDSLINE, CANCERLIT y EMBASE, así como la revisión manual de la bibliografía de estudios relevantes publicados, de revisiones anteriores, la evaluación de la base de datos Cochrane Collaboration y la consulta con expertos. La estrategia de búsqueda debe describirse claramente, de modo que otros investigadores puedan reproducirla.

---

1. Pregunta de la investigación clara
2. Identificación exhaustiva y sin sesgo de los estudios realizados
3. Definición clara de los criterios de inclusión y exclusión
4. Resumen uniforme y sin sesgo de las características y hallazgos de cada estudio
5. Presentación clara y uniforme de los datos de los estudios individuales
6. Cálculo de un estimador resumen del efecto y del intervalo de confianza basado en los hallazgos de todos los estudios elegibles, cuando proceda
7. Valoración de la heterogeneidad de los hallazgos de los estudios individuales
8. Valoración de posible sesgo de publicación
9. Análisis de subgrupos y análisis de sensibilidad

Criterios para incluir y excluir estudios

El protocolo de una revisión sistemática debe proporcionar una buena justificación para incluir y excluir estudios, y estos criterios deben establecerse a priori (tabla 13-2). Una vez establecidos estos criterios, debe revisarse independientemente por dos o más investigadores la idoneidad de cada uno de los estudios potencialmente elegible, y los desacuerdos debe resolverlos otro revisor o se resolverán por consenso. Cuando se determine la elegibilidad, puede que sea mejor que los revisores desconozcan la fecha, la revista, los autores y los resultados de los ensayos.

Las revisiones sistemáticas publicadas deben enumerar estudios cuya inclusión se consideró, y la razón concreta para excluir un estudio. Por ejemplo, si se identifican 30 estudios potencialmente elegibles, deben referenciarse totalmente y deberá darse una razón por cada exclusión que se haga.

Recopilación de datos de los estudios elegibles

Los datos deben obtenerse de cada estudio de un modo uniforme y no sesgado. Generalmente, esto lo realizan independientemente dos o más compendiadores, que usan formularios prediseñados (tabla 13-3). Los formularios de resumen de datos deben incluir todos los datos que aparecerán posteriormente en el texto, las tablas o las figuras que describen los estudios incluidos en la revisión sistemática, o en las tablas o las figuras que presentan las respuestas. Cuando los dos compendiadores no estén de acuerdo, un tercero puede marcar la diferencia, o puede usarse un proceso de consenso. El proceso para resumir los datos de los estudios para la revisión sistemática debe describirse claramente en el manuscrito.

Los artículos publicados de algunos estudios que podrían elegirse para su inclusión en una revisión sistemática pueden no contener información importante, como características del diseño, cálculos de riesgo y desviaciones típicas. A menudo resulta difícil decir si algunas características del diseño, como el enmascaramiento, no se realizaron o no estaban descritas en la publicación. El revisor puede calcular, a veces, riesgos relativos e intervalos de confianza a partir de datos no elaborados presentados en *ensayos* aleatorizados, pero no suele aceptarse el cálculo de estimadores del riesgo y de intervalos de confianza basándose en datos no elaborados de estudios de *observación*, porque no hay suficiente información para ajustar las variables de confusión. Debe realizarse un esfuerzo por contactar con los autores para recuperar información importante que no esté incluida en la descripción publicada de un estudio. Si esta información necesaria no puede calcularse ni obtenerse, suelen excluirse los hallazgos del estudio.

CRITERIOS	**EJEMPLO: ÁCIDOS GRASOS Ω-3 Y EPISODIOS CARDIOVASCULARES***
1. Período durante el que se publicaron los estudios	Estudios publicados antes de agosto de 2 012
2. Diseño del estudio	Estudios aleatorizados y controlados realizados en contextos de prevención primaria o secundaria de enfermedades cardiovasculares
3. Población del estudio	Estudios de adultos aleatorizados a ácidos grasos ω-3 o al testigo
4. Intervención o factor de riesgo	Administración de ácidos grasos ω-3, con la dieta o con suplementos, en cualquier dosis, administrados durante al menos 1 año
5. Grupos testigos aceptables	Dieta o suplementos sin ácidos grasos ω-3
6. Otros requisitos del diseño del estudio (p.ej., enmascaramiento en ensayos clínicos o control de posibles factores de confusión específicos en estudios de observación)	Ninguno
7. Variables de respuesta aceptables	Mortalidad por todas las causas, muerte cardíaca, muerte súbita, infarto de miocardio y accidente cerebrovascular
8. Pérdida durante el seguimiento máxima aceptable	No declarada
9. Duración mínima aceptable del seguimiento	No declarada

* Este ejemplo de cómo se utilizan estos criterios procede de un metaanálisis publicado en el que no se observó ningún efecto de los ácidos grasos ω-3 para la prevención de episodios de enfermedad cardiovascular (24).

---

1. Criterios de elegibilidad (¿en qué medida el estudio cumple los criterios de elegibilidad establecidos?)
2. Características del diseño (diseño del estudio, grupo testigo, enmascaramiento, control de los factores de confusión, etc.)
3. Características y número de los participantes en cada uno de los grupos del estudio (datos demográficos, gravedad de la enfermedad, etc.)
4. Intervención (en los ensayos) o factores de riesgo (en los estudios de observación)
    - Para las intervenciones: dosis, duración del tratamiento, etc.
    - Para los estudios de observación: tipo y nivel del factor de riesgo, etc.
5. Criterio de valoración principal, criterios de valoración secundarios y criterios de valoración en subgrupos preestablecidos
6. Elementos para permitir la evaluación de la calidad de los estudios incluidos (aleatorización, enmascaramiento, cumplimiento, pérdida durante el seguimiento, control de los factores de confusión, etc.)

Cómo presentar claramente los hallazgos

Las revisiones sistemáticas suelen incluir tres tipos de información. En primer lugar, las características importantes de cada estudio incluido en la revisión sistemática se presentan en tablas, que a menudo incluyen características de la población en estudio, el tamaño de la muestra, la cantidad o la incidencia de las variables de respuesta, la duración del seguimiento y los métodos usados en el estudio. En segundo lugar, la revisión ofrece los hallazgos analíticos de los estudios individuales (riesgo relativo, razón de posibilidades, diferencia de riesgo e intervalos de confianza o valores de *p*) en una tabla o figura. Finalmente, si no hay heterogeneidad importante (v. más adelante), el metaanálisis presenta estimadores resumen e intervalos de confianza basados en los hallazgos de todos los estudios incluidos, así como análisis de sensibilidad y de subgrupos.

Los estimadores resumen del efecto representan una respuesta principal del metaanálisis, pero deben presentarse en el contexto de toda la información extraída de los estudios individuales. Las características y hallazgos de los estudios individuales incluidos en la revisión sistemática deben disponerse claramente en tablas y figuras, de modo que el lector pueda formarse opiniones que no solo dependan de los estimadores resumen estadísticos.

Metaanálisis: estadísticos para revisiones sistemáticas

- Estimador resumen del efecto e intervalo de confianza. Cuando se han identificado todos los estudios realizados, se han elegido los que cumplen los criterios de inclusión y exclusión, y se han obtenido datos de cada estudio, generalmente se calcula un intervalo de confianza y un estimador resumen (riesgo relativo resumen, razón de posibilidades resumen, diferencia de riesgos resumen, etc.). El efecto resumen es, esencialmente, un efecto promedio ponderado por la inversa de la varianza del criterio de valoración de cada uno de los estudios. En el apéndice 13 se comentan los métodos para calcular el estimador resumen del efecto y el intervalo de confianza. Los que no estén interesados en los detalles del cálculo de los estimadores medios ponderados a partir de múltiples estudios deben, al menos, conocer que diferentes métodos pueden dar resultados distintos. Por ejemplo, recientes metaanálisis de la eficacia de los preservativos para evitar la transmisión heterosexual del VIH han proporcionado estimadores resumen que oscilan entre una disminución del 80 % al 94 % de la incidencia de transmisión, aunque se basan en los resultados de series de estudios casi idénticas (18 , 19).
- Heterogeneidad. No es adecuado combinar los resultados de varios estudios si estos difieren en aspectos con importancia clínica, como la intervención, la respuesta, la situación testigo, el enmascaramiento, etc. Incluso si los métodos usados en los estudios parecen similares, el hecho de que los resultados varíen mucho sugiere que había diferencias en algún aspecto importante en los estudios individuales. Esta variabilidad de los hallazgos de los estudios individuales es lo que se denomina heterogeneidad (y los hallazgos del estudio se denominan heterogéneos); si existe poca variabilidad, se dice que los resultados del estudio son homogéneos.

¿Cómo puede decidir el investigador si los métodos y los hallazgos son lo suficientemente similares como para combinarlos en

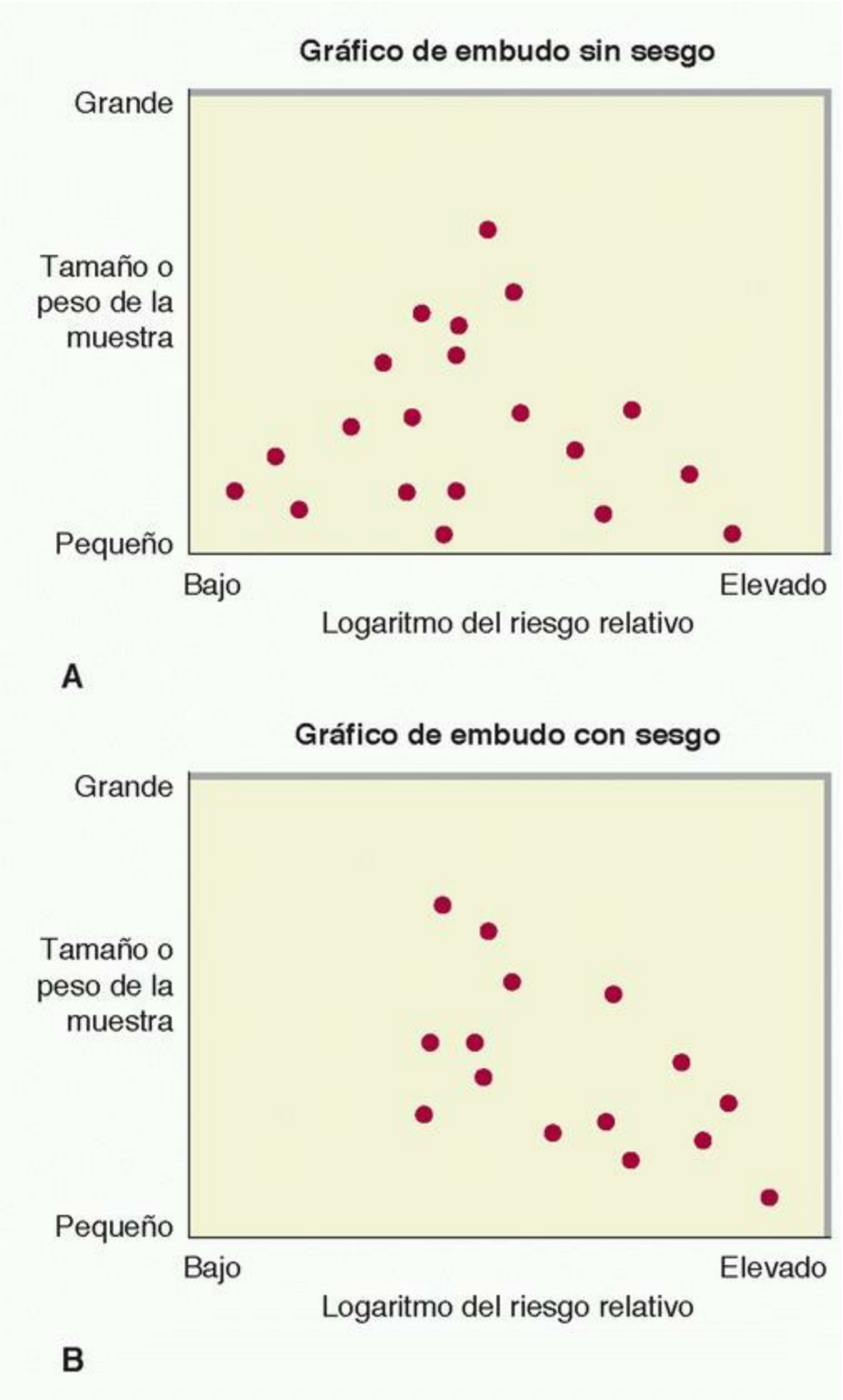
estimadores resumen? En primer lugar, puede revisar los estudios individuales para determinar si hay diferencias importantes en el diseño del estudio, las poblaciones del estudio, la intervención o la respuesta. Luego, puede examinar los resultados de los estudios concretos. Si algunos ensayos comunican un importante efecto beneficioso de una intervención y otros comunican un considerable perjuicio, existe, claramente, heterogeneidad. A veces, es difícil decidir si hay heterogeneidad. Por ejemplo, si un ensayo comunica una reducción del 50 % del riesgo para una intervención específica, pero otro solo comunica una reducción del 30 %, ¿existe heterogeneidad? Se han desarrollado métodos estadísticos (pruebas de homogeneidad) para ayudar a responder a esta pregunta (v. apéndice 13), pero, finalmente, se debe aplicar el sentido común cuando se evalúe la heterogeneidad. Todas las revisiones sistemáticas publicadas deben incluir algún comentario sobre heterogeneidad y su efecto sobre los estimadores resumen.

Valoración del sesgo de publicación

Se produce sesgo de publicación cuando los estudios publicados no son representativos de todos los estudios que se han realizado, generalmente porque los resultados positivos tienden a enviarse y publicarse con mayor frecuencia que los negativos. Hay dos modos de abordar el sesgo de publicación. Pueden identificarse los estudios no publicados para incluir los resultados en un estimador resumen. Los resultados no publicados pueden identificarse consultando a los investigadores y revisando resúmenes, presentaciones de congresos y tesis doctorales. Los resultados de estudios no publicados pueden incluirse con los de estudios publicados en el estimador resumen general, o los análisis de sensibilidad pueden determinar si la adición de estos resultados no publicados cambia sustancialmente el estimador resumen determinado a partir de resultados publicados. Sin embargo, la inclusión en una revisión sistemática de resultados no publicados es problemática por varios motivos. A menudo es difícil identificar los estudios no publicados, e incluso es más difícil extraer los datos precisos. Frecuentemente se dispone de información inadecuada para determinar si el estudio cumple los criterios de inclusión para la revisión sistemática o para evaluar la calidad de los métodos (que, al no tener el rigor de la revisión externa, puede ser inferior). Por estos motivos, a menudo no se incluyen datos no publicados en metaanálisis.

Por otro lado, puede calcularse la magnitud del posible sesgo de publicación, y esta información puede usarse para moderar las conclusiones de la revisión sistemática. Existe sesgo de publicación cuando estudios no publicados tienen hallazgos diferentes a los de los estudios publicados. Es más probable que los no publicados sean pequeños (los estudios grandes suelen publicarse, independientemente de los hallazgos) y que no hayan encontrado ninguna asociación entre el factor de riesgo o la intervención y la respuesta (los estudios notablemente positivos suelen publicarse, aunque sean pequeños). Si no hay sesgo de publicación, no debe haber ninguna asociación entre el tamaño del estudio (o la varianza de la variable de respuesta) y los hallazgos. El grado de esta asociación se mide, a menudo, usando la τde Kendall, un coeficiente de correlación. Una relación intensa o estadísticamente significativa entre la respuesta del estudio y el tamaño de la muestra sugiere sesgo de publicación. Si no existe este sesgo, una representación gráfica del tamaño de la muestra del estudio frente a la respuesta (p. ej., logaritmo del riesgo relativo) debe tener forma de campana o forma de embudo, con la punta cerca del estimador resumen del efecto.

El gráfico de embudo de la figura 13-1A sugiere que existe poco sesgo de publicación, porque se publicaron estudios pequeños con hallazgos tanto positivos como negativos. Por otro lado, el gráfico de la figura 13-1B sugiere que existe sesgo de publicación, porque la distribución aparece truncada en la esquina que debe contener estudios pequeños y con resultados negativos.



▪FIGURA 13-1. A) Gráfico de embudo que no indica un sesgo de publicación, porque hay estudios con un espectro de tamaños de muestra grandes y pequeños, y algunos estudios más pequeños comunican riesgos relativos bajos. B) Gráfico de embudo que sugiere sesgo de publicación, porque pocos de los estudios más pequeños comunican riesgos relativos bajos.
Herramientas de imágenes

Cuando es probable que exista un importante sesgo de publicación, no deben calcularse los estimadores resumen, o deben interpretarse con precaución. Todas las revisiones sistemáticas publicadas deben incluir algún comentario sobre el posible sesgo de publicación y su efecto sobre los estimadores resumen.

Análisis de subgrupos y de sensibilidad

Se pueden realizar análisis de subgrupos usando datos de todos o de algunos de los subgrupos de los estudios incluidos en la revisión sistemática. Por ejemplo, en una revisión sistemática del efecto del tratamiento posmenopáusico con estrógenos sobre el riesgo de sufrir cáncer endometrial, algunos de los estudios presentaban los resultados según la duración del uso de los estrógenos. Los análisis por subgrupos de los resultados de los estudios que proporcionaron esta información demostraron que el uso más prolongado se asociaba a un mayor riesgo de sufrir cáncer (20).

Los análisis de sensibilidad indican la «sensibilidad» de los hallazgos del metaanálisis para ciertas decisiones sobre el diseño de la revisión sistemática o la inclusión de determinados estudios. Por ejemplo, si los autores deciden incluir estudios con un diseño o con métodos algo diferentes en la revisión sistemática, los hallazgos se refuerzan si los resultados resumen son similares tanto si se incluyen como si no se incluyen los estudios cuestionables. Las revisiones sistemáticas deben, generalmente, incluir análisis de sensibilidad si alguna de las decisiones del diseño parece cuestionable o arbitraria.

Los metaanálisis pueden incrementar la potencia para responder una pregunta de la investigación, aunque tienen la desventaja de que no incluyen los datos a nivel individual para permitir el ajuste de las posibles variables de confusión o para realizar análisis de subgrupos individuales. En algunos casos puede ser posible obtener los datos de los individuos de los correspondientes estudios individuales y realizar análisis conjuntos. En estos casos se pueden utilizar los datos combinados de los estudios individuales para ajustar los factores de confusión o evaluar los efectos en diversos subgrupos, igual que se realizaría en un único estudio extenso. Por ejemplo, el Early Breast Cancer Trialists Collaborative Group combinó datos individuales de 123 estudios aleatorizados para evaluar la eficacia de diferentes regímenes quimioterápicos en el cáncer de mama temprano (21). No obstante, generalmente es difícil obtener datos a nivel individual de los correspondientes estudios, y es infrecuente que en estos estudios se hayan medido las variables de manera que sean suficientemente similares como para poder combinarlas en un único conjunto de datos.

Entra basura, sale basura

El mayor inconveniente de una revisión sistemática es que puede producir un estimador resumen de aspecto fiable basado en los resultados de estudios individuales de escasa calidad. En los metaanálisis se utilizan diversos abordajes para evaluar la calidad de los diferentes diseños de estudio, aunque el proceso de la evaluación de la calidad es complejo y problemático. Preferimos confiar en criterios relativamente estrictos para un buen diseño del estudio al establecer los criterios de inclusión. Si los estudios individuales que se resumen en una revisión sistemática son de escasa calidad, ningún análisis cuidadoso puede evitar que el estimador resumen no sea fiable. Una cuestión especial de este problema se encuentra en las revisiones sistemáticas de datos de observación. Si los resultados de estos estudios no se ajustan teniendo en cuenta las posibles variables de confusión, los resultados del metaanálisis tampoco estarán ajustados y podrá existir confusión.

Volver al principio

▪ RESUMEN

En este capítulo se describen tres abordajes para hacer un uso creativo de los datos y muestras existentes, una forma eficaz que tienen los nuevos investigadores con pocos recursos de adquirir una experiencia valiosa y una publicación rápida.

Análisis de datos secundarios

- Este abordaje de la utilización de conjuntos de datos existentes tiene la ventaja de reducir mucho el tiempo y el coste de realizar la investigación, y la desventaja de que permite un control escaso o nulo de la población del estudio, el diseño y las mediciones.
- Las fuentes de datos de los análisis secundarios incluyen proyectos de investigación ya existentes, historias médicas electrónicas, bases de datos administrativas y bases de datos públicas, como registros de tumores, registros de certificados de defunción y estudios regionales, como el estudio NHANES.
- Las grandes bases de datos comunitarias son útiles para estudiar la eficacia (los efectos en la práctica clínica real de una intervención en diversas comunidades), para evaluar las tasas de utilización y la variación regional, y para descubrir acontecimientos adversos infrecuentes.
- Los estudios de asociaciones basados en datos conjuntos se denominan estudios ecológicos; pueden aportar información útil, aunque están sometidos a sesgos especiales, denominados falacias ecológicas.

Estudios auxiliares

- Un estudio auxiliar es un análisis de datos secundarios en el que el investigador realiza una o más nuevas mediciones para

responder a una nueva pregunta de la investigación con un coste y un esfuerzo relativamente pequeños.

- Pueden encontrarse buenas oportunidades para hacer estudios auxiliares en estudios de cohortes o ensayos clínicos que incluyen la variable predictiva o la variable de respuesta para la nueva pregunta de la investigación.
- Las muestras de suero, ADN, imágenes, etc., almacenadas ofrecen la oportunidad de realizar diseños de casos y testigos anidados.
- La mayor parte de los grandes estudios tienen políticas escritas que permiten que los investigadores (incluso científicos externos) propongan y lleven a cabo análisis secundarios de los datos y estudios auxiliares

Revisiones sistemáticas

- Una buena revisión sistemática, como cualquier otro estudio, necesita un protocolo por escrito completo antes de iniciar el estudio; dicho protocolo debe incluir la pregunta del estudio, los métodos para identificar todos los estudios elegibles, los métodos para extraer datos de los estudios y los métodos estadísticos.
- Los aspectos estadísticos de la combinación de estudios sobre un tema, denominada metaanálisis, incluyen el estimador resumen del efecto y el intervalo de confianza, pruebas para evaluar la heterogeneidad y el posible sesgo de publicación, y análisis de subgrupos y de sensibilidad.
- Las características y los hallazgos de estudios individuales deben mostrarse claramente en tablas y figuras, de modo que el lector pueda formarse opiniones que no dependan tan solo de los estimadores resumen estadísticos.
- Un importante problema es la evaluación de la calidad de los estudios de una revisión sistemática, que puede influir mucho en los hallazgos de la revisión.

Volver al principio

APÉNDICE 13Métodos estadísticos para el metaanálisis
NA

▪ RESUMEN DEL EFECTO E INTERVALOS DE CONFIANZA

El principal objetivo del metaanálisis es calcular un estimador resumen del efecto y un intervalo de confianza. Un modo intuitivo de hacerlo es multiplicar cada variable de respuesta del estudio, como el riesgo relativo (una estimación del efecto), por el tamaño de la muestra (un peso que refleja la exactitud del riesgo relativo), sumar estos productos y dividir por la suma de los pesos. En la práctica real, se usa como el peso para cada uno de los estudios el inverso de la varianza del estimador del efecto para cada estudio individual (1/varianza). El inverso de la varianza es una estimación de la precisión del estimador del efecto mejor que el tamaño de la muestra, porque tiene en cuenta el número de respuestas y su distribución. El estimador del efecto medio ponderado se calcula multiplicando cada valor relativo del estudio (1/varianza) por el logaritmo del riesgo relativo (o cualquier otro estimado de riesgo, como el logaritmo del cociente de posibilidades, la diferencia de riesgos, etc.), sumando estos productos y dividiéndolo por la suma de los pesos relativos. Los estudios pequeños suelen dar una gran varianza (y un intervalo de confianza amplio alrededor del estimador del riesgo), y los grandes estudios dan una varianza pequeña (y un intervalo de confianza estrecho alrededor del estimador del riesgo). Por lo tanto, en un metaanálisis, los estudios extensos tienen mucho peso relativo (1/varianza pequeña) y los pequeños estudios tienen un peso relativo bajo (1/varianza grande).

Para determinar si el estimador resumen del efecto es estadísticamente significativo, se calcula la variabilidad del estimador resumen del efecto. Hay varias fórmulas para calcular la varianza de los estimadores resumen del riesgo (22 , 23). La mayor parte usa algo que se aproxima al inverso de la suma de los pesos relativos de los estudios individuales ($1/\Sigma$ peso$_i$). La varianza del estimador resumen se usa para calcular el intervalo de confianza del 95 % alrededor del estimador resumen ($\pm$ 1,96 $\times$ varianza$^{1/2}$).

▪ MODELOS DE EFECTOS ALEATORIOS Y DE EFECTOS FIJOS

Se dispone de múltiples métodos estadísticos para calcular un estimador resumen (22, 23). La elección del método estadístico suele depender del tipo de respuesta (riesgo relativo, cociente de posibilidades, diferencia de riesgos, etc.). Además del modelo estadístico, el investigador debe también escoger el uso de un modelo de efectos fijos o de efectos aleatorios. El modelo de efectos fijos simplemente calcula la varianza de un estimador resumen ponderado basado en el inverso de la suma de los pesos de cada estudio individual. El modelo de efectos aleatorios suma la varianza al resumen del efecto en proporción a la variabilidad de los resultados de los estudios

individuales. Los estimadores resumen del efecto suelen ser similares usando uno u otro modelo, pero la varianza del resumen del efecto es mayor con el modelo de efectos aleatorios, en la medida en que difieren los resultados de los estudios individuales y el intervalo de confianza alrededor del resumen del efecto es, en consecuencia, mayor, de modo que los resultados resumen tienen menos probabilidad de ser estadísticamente significativos. Muchas revistas piden actualmente a los autores que usen un modelo de efectos aleatorios, porque se considera que es «conservador» (es decir, es menos probable que encuentre un efecto estadísticamente significativo si no existe). En los metaanálisis se debe establecer claramente si se usa un modelo de efectos fijos o de efectos aleatorios.

La simple utilización de un modelo de efectos aleatorios no evita el problema de la heterogeneidad. Si los estudios identificados mediante una revisión sistemática son claramente heterogéneos, no debe calcularse un estimador resumen.

▪ PRUEBAS ESTADÍSTICAS DE HOMOGENEIDAD

Las pruebas de homogeneidad suponen que los hallazgos de los ensayos individuales son los mismos (la hipótesis nula), y se usa una prueba estadística (prueba de homogeneidad) para determinar si los datos (los hallazgos de los estudios individuales) refutan esta hipótesis. Suele utilizarse la prueba de la χ^2 (22). Si los datos apoyan la hipótesis nula (valor de $p \geq 0{,}10$), el investigador acepta que los estudios son homogéneos. Si los datos no apoyan la hipótesis (valor de $p < 0{,}10$), rechaza la hipótesis nula y supone que los hallazgos del estudio son heterogéneos. En otras palabras, existen diferencias significativas en las poblaciones estudiadas, en la naturaleza de las variables de respuesta o en los resultados del estudio.

Todos los metaanálisis deben describir las pruebas de homogeneidad con un valor de *p*. Estas pruebas no son muy potentes, y es difícil rechazar la hipótesis nula y demostrar heterogeneidad cuando el tamaño de la muestra (número de personas estudiadas) es pequeño. Por este motivo, un valor de *p* de 0,10 en lugar de 0,05 es el que habitualmente se usa como valor de corte. Si existe una notable heterogeneidad, no es adecuado combinar los resultados de los ensayos en un solo estimador resumen.

BIBLIOGRAFÍA

1. Yaffe K, Barnes D, Nevitt M, et al. A prospective study of physical activity and cognitive decline in elderly women: women who walk. Arch Intern Med 2001;161:1703-1708. Citado aquí

2. Kerlikowske K, Miglioretti D, Buist D, et al. Declines in invasive breast cancer and use of postmenopausal hormone therapy in a screening mammography population. J Natl Cancer Inst. 2007;99:1335-1339. Citado aquí

3. Canner PL. Mortality in CDP patients during a nine-year post-treatment period. J Am Coll Cardiol 1986;8:1243-1255. Citado aquí

4. Looker AC, Johnston CC Jr., Wahner HW, et al. Prevalence of low femoral bone density in older U.S. women from NHANES III. J Bone Miner Res 1995;10:796-802. Citado aquí

5. Looker AC, Melton LJ, Harris TB, et al. Prevalence and trends in low femur bone density among older US adults: NHANES 2005-2006 compared with NHANES III. J Bone Miner Res 2010;25:64-71. Citado aquí

6. Gurwitz JH, Gore JM, Goldberg RJ, et al. Risk for intracranial hemorrhage after tissue plasminogen activator treatment for acute myocardial infarction. Participants in the National Registry of Myocardial Infarction 2. Ann Intern Med 1998;129:597-604. Citado aquí

7. Weaver WD, Simes RJ, Betriu A, et al. Comparison of primary coronary angioplasty and intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review. JAMA 1997;278:2093-2098; published erratum appears in JAMA 1998;279:876. Citado aquí

8. Every NR, Parsons LS, Hlatky M, et al. A comparison of thrombolytic therapy with primary coronary angioplasty for acute myocardial infarction. Myocardial infarction triage and intervention investigators. N Engl J Med 1996;335:1253-1260. Citado aquí

9. Tiefenbrunn AJ, Chandra NC, French WJ, et al. Clinical experience with primary percutaneous transluminal coronary angioplasty compared with alteplase (recombinant tissue-type plasminogen activator) in patients with acute myocardial infarction: a report from the Second National Registry of Myocardial Infarction (NRMI-2). J Am Coll Cardiol 1998;31:1240-1245. Citado aquí

10. Cassard SD, Quigley HA, Gower EW, et al. Regional variations and trends in the prevalence of diagnosed glaucoma in the Medicare population. Ophthalmology 2012;119:1342-1351. Citado aquí

11. Hearst N, Newman TB, Hulley SB. Delayed effects of the military draft on mortality: a randomized natural experiment. N Engl J Med 1986;314:620-624. Citado aquí

12. Shlipak M, Simon J, Vittinghoff E, et al. Estrogen and progestin, lipoprotein (a), and the risk of recurrent coronary heart disease events after menopause. JAMA 2000;283:1845-1852. Citado aquí

13. Sawaya GF, Grady D, Kerlikowske K, et al. The positive predictive value of cervical smears in previously screened postmenopausal women: the Heart and Estrogen/Progestin Replacement Study (HERS). Ann Intern Med 2000;133:942-950. Citado aquí

14. Grady D, Brown J, Vittinghoff E, et al. Postmenopausal hormones and incontinence: the Heart and Estrogen/Progestin Replacement Study. Obstet Gynecol 2001;97:116-120. Citado aquí

15. Grady D, Yaffe K, Kristof M, et al. Effect of postmenopausal hormone therapy on cognitive function: the Heart and Estrogen/Progestin Replacement Study. Am J Med 2002;113:543-548. Citado aquí

16. Herrington DM, Vittinghoff E, Howard TD, et al. Factor V Leiden, hormone replacement therapy, and risk of venous thromboembolic events in women with coronary disease. Arterioscler Thromb Vasc Biol 2002;22:1012-1017. Citado aquí

17. Oler A, Whooley M, Oler J, et al. Heparin plus aspirin reduces the risk of myocardial infarction or death in patients with unstable angina. JAMA 1996;276:811-815. Citado aquí

18. Pinkerton SD, Abramson PR. Effectiveness of condoms in preventing HIV transmission. Soc Sci Med 1997;44: 1303-1312. Citado aquí

19. Weller S, Davis K. Condom effectiveness in reducing heterosexual HIV transmission. Cochrane Database Syst Rev 2002; (1):CD003255. Citado aquí

20. Grady D, Gebretsadik T, Kerlikowske K, et al. Hormone replacement therapy and endometrial cancer risk: a meta-analysis. Obstet Gynecol 1995;85:304-313. Citado aquí

21. Peto R, Davies C, Godwin J, et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet 2012;379:432-441. Citado aquí

22. Petitti DB. Meta-analysis, decision analysis and cost effectiveness analysis: methods for quantitative synthesis in medicine, 2nd ed. New York: Oxford University Press, 2000. Citado aquí

23. Cooper H, Hedges LV. The handbook of research synthesis. New York: Russell Sage Foundation, 1994. Citado aquí

24. Rizos EC, Ntzani EE, Bika E, et al. Association between omega-3 fatty acid supplementation and risk of major cardiovascular disease events. JAMA 2012;308:1024-1033.

CAPÍTULO 14 Abordaje de cuestiones éticas

Bernard Lo
Deborah G. Grady

La investigación con participantes humanos conlleva problemas éticos, porque las personas aceptan molestias y riesgos para el progreso del conocimiento científico y para beneficiar a otros. El público, que participa en la investigación y la financia, debe poder confiar en que la investigación sigue criterios éticos elevados.

En este capítulo comenzamos con la historia de la supervisión de la investigación, y después revisamos los principios éticos y las leyes federales que guían la investigación con participantes humanos, especialmente los requisitos de autorización por el Comité ético de la investigación clínica (CEIC) y el consentimiento informado. Finalmente, revisamos aspectos relacionados con la conducta científica impropia, la autoría, los conflictos de intereses y los aspectos éticos en tipos de investigación específicos.

▪ HISTORIA DE LAS REGULACIONES SOBRE LA INVESTIGACIÓN CLÍNICA

Las regulaciones y directrices actuales para la investigación clínica han respondido a los abusos cometidos, como la «investigación» de los médicos nazis durante la Segunda Guerra Mundial, la investigación en los Estados Unidos con prisioneros, personas internadas en centros residenciales a largo plazo y otras poblaciones vulnerables, y el estudio de Tuskegee (caso 14-1).

CASO 14-1. El estudio de Tuskegee

En 1932, las agencias gubernamentales estadounidenses iniciaron el estudio de Tuskegee para documentar la evolución natural y los efectos crónicos de la sífilis no tratada. Los pacientes eran hombres afroamericanos de zonas rurales de Alabama con un nivel educativo y económico bajo. Recibieron alimentos, una asistencia médica básica y un seguro de entierro. Los investigadores dijeron falsamente a los pacientes que estaban recibiendo tratamiento para la sífilis, por ejemplo, diciéndoles que las punciones lumbares que se realizaban con fines de investigación eran «tratamientos especiales gratuitos». Cuando se empezó a disponer de antibióticos para la sífilis durante la Segunda Guerra Mundial, y cuando posteriormente se recomendaron como medida de salud pública, los investigadores tomaron medidas para evitar que los pacientes recibieran el tratamiento. En respuesta al estudio de Tuskegee, en 1974, el gobierno federal publicó regulaciones sobre la investigación en pacientes humanos, que precisaba el consentimiento informado de los participantes y la revisión por los CEIC de la investigación en participantes humanos patrocinada por el gobierno federal. En 1997, el presidente Clinton pidió disculpas formales por el estudio de Tuskegee.

Volver al principio

▪ PRINCIPIOS ÉTICOS

Los fallos éticos del estudio de Tuskegee y de otros diversos estudios inspiraron las regulaciones actuales para la protección de los participantes en estudios de investigación. Se articularon tres principios éticos, que habían sido violados en estos estudios, para guiar la investigación con participantes humanos (2). Primero, el reconocimiento de que todas las personas tienen derecho a tomar sus propias decisiones sobre su participación en investigaciones; el principio del respeto por las personas exige que los investigadores obtengan el consentimiento informado y voluntario de los participantes en la investigación, les permitan suspender su participación en la investigación en cualquier momento y protejan a los participantes que tengan mermada su capacidad de tomar decisiones.

Segundo, el principio de beneficencia exige que el consenso científico que se obtenga del estudio deba superar las molestias y los riesgos que experimenten los participantes en la investigación, y que se mantengan al mínimo dichos riesgos. Los riesgos incluyen tantos daños físicos por las intervenciones de la investigación, como perjuicios psicosociales, vulneraciones de la confidencialidad, estigma y discriminación. Los riesgos de los participantes en el estudio pueden reducirse, por ejemplo, realizando pruebas de selección a los posibles participantes para excluir a los que tengan probabilidad de sufrir daños, garantizando la confidencialidad y siguiendo a los pacientes para detectar la aparición de efectos adversos.

Tercero, el principio de justicia precisa que los beneficios y las cargas de la investigación se distribuyan objetivamente. No se debe incluir selectivamente como participantes a poblaciones desfavorecidas y vulnerables, como personas con rentas bajas, nivel educativo bajo, escaso acceso a la asistencia sanitaria o alteración de la capacidad de tomar decisiones, si otras poblaciones también serían adecuadas para abordar las preguntas de la investigación. El estudio de grupos vulnerables, principalmente debido a la facilidad del acceso, la colaboración y el seguimiento, se aprovecha injustamente de ellos.

La justicia también precisa un acceso equitativo a los beneficios de la investigación. Tradicionalmente, se ha contemplado la investigación clínica como algo arriesgado, y se ha pensado en los posibles participantes como en conejillos de Indias a los que se debía proteger de

intervenciones peligrosas que les proporcionarían pocos o ningún beneficio personal. Sin embargo, cada vez más se contempla la investigación clínica como un modo de poder acceder a nuevos tratamientos para afecciones, como la infección por el VIH y el cáncer. Los pacientes que buscan prometedores fármacos nuevos para afecciones mortales desean tener un mayor acceso a la investigación clínica, no una mayor protección, y este acceso debe estar disponible independientemente de los ingresos, el seguro o el nivel educativo. Históricamente, los niños, las mujeres y los miembros de minorías étnicas han estado representados de manera insuficiente en la investigación clínica, lo que ha llevado a una débil base de datos científicos y a la posibilidad de un tratamiento clínico subóptimo. El principio de justicia precisa que se incluya a estos grupos en los estudios de investigación. En las investigaciones clínicas financiadas por los NIH debe haber una representación adecuada de niños, mujeres y miembros de minorías étnicas, o se debe justificar por qué estos grupos podrían estar representados de forma insuficiente.

Volver al principio

▪ LEGISLACIÓN FEDERAL PARA LA INVESTIGACIÓN CON PARTICIPANTES HUMANOS

La legislación federal se aplica a toda la investigación financiada con fondos federales y a la investigación que se enviará a la U.S. Food and Drug Administration (FDA) como respaldo de la solicitud de comercialización de un nuevo fármaco o dispositivo. Además, las universidades exigen que toda la investigación en participantes humanos realizada por el personal docente y por el resto del personal cumpla las regulaciones fundamentales sobre el consentimiento informado y la revisión por el CEIC; esto también afecta a la investigación financiada privadamente y a la que se realiza en un centro externo. Aunque las normas se refieren a «sujetos» humanos, algunos autores prefieren el término «participantes», porque insisten en que las personas son participantes activos en la investigación y no sujetos en los que se realizan experimentos.

Es importante que se comprendan varias definiciones de estas normas:

- La investigación es una «búsqueda sistemática diseñada para desarrollar o contribuir al conocimiento generalizable» (3). No se considera que sea investigación la asistencia clínica de eficacia no demostrada que se dirige a beneficiar al paciente individual y no a la publicación. Se podrían tratar como investigación algunos proyectos para la mejora de la calidad, aunque la mayoría cumplen los criterios de exención, que se analizarán después.

- Los participantes humanos son personas vivas de las que un investigador obtiene «datos a través de una intervención o interacción con la persona» o «información privada identificable».

- La información privada comprende: *1)* información que una persona puede esperar razonablemente que no se está observando ni registrando, y *2)* información que se ha proporcionado con fines específicos y que «la persona puede esperar razonablemente que no se hará pública (p. ej., una historia clínica)». La información es identificable si «la identidad de la persona se verifica o puede verificarse por parte del investigador».

- Los datos de la investigación codificados no se pueden identificar si la clave que relaciona los datos con los participantes se destruye antes de que empiece la investigación o si los investigadores no tienen acceso a la clave.

Las leyes federales sobre la protección de los sujetos humanos (*Federal Regulations on the Protection of Human Subjects*) están disponibles en la página web de la oficina para la protección de la investigación en seres humanos (Office for Human Research Protections) (3). Los investigadores que tengan dudas sobre estas leyes federales deben consultar con su CEIC. Estas leyes federales proporcionan dos importantes protecciones a los participantes humanos: la aprobación por los CEIC y el consentimiento informado.

Aprobación por el CEIC

Las leyes federales precisan que la investigación con participantes humanos sea aprobada por un CEIC. La misión de los CEIC es asegurar que la investigación sea éticamente aceptable y que se protegen el bienestar y los derechos de los participantes en la investigación. Aunque la mayoría de los miembros de los CEIC son investigadores, los CEIC deben incluir miembros de la comunidad y personas que conozcan bien los aspectos legales y éticos que conciernen a la investigación.

Cuando aprueba un estudio de investigación, el CEIC debe determinar que (3):

- Se han reducido al mínimo los riesgos para los participantes.

- Los riesgos son razonables en relación con los beneficios previstos y la importancia de la información que se espera obtener.

- La selección de los participantes es equitativa.

- Se obtendrá el consentimiento informado de los participantes o de sus representantes legalmente autorizados.
- Se mantiene adecuadamente la confidencialidad.

El sistema de CEIC está descentralizado. Cada CEIC local aplica las leyes federales usando sus propios formularios, procedimientos y directrices, y no se recurre a un organismo superior. Por ello, un protocolo de un estudio multicéntrico podría ser autorizado por un CEIC pero no por otro. Generalmente estas diferencias pueden resolverse mediante conversaciones o modificaciones del protocolo.

Se ha criticado a los CEIC y a las leyes federales por varios motivos (4 , 5). Pueden hacer un hincapié excesivo en los formularios del consentimiento, no examinar el diseño de la investigación y no tener en consideración adecuadamente el valor científico de la investigación. Aunque los CEIC deben estudiar todas las revisiones de protocolos y realizar un seguimiento de los acontecimientos adversos, no suelen comprobar si la investigación se realizó realmente según los protocolos aprobados. Muchos CEIC carecen de recursos y experiencia para cumplir adecuadamente su misión de proteger a los participantes en las investigaciones. Por estos motivos, debe considerarse que las leyes federales y la aprobación por el CEIC son solo una norma ética mínima para la investigación. En último término, la sensatez y el carácter del investigador constituyen el elemento más esencial para asegurar que la investigación sea éticamente aceptable.

Excepciones a la revisión completa por el CEIC

- La mayoría de los estudios de investigación en los que se utilizan encuestas y entrevistas, además de los análisis secundarios de registros y muestras existentes sin identificación, pueden quedar exentos de la revisión por el CEIC (tabla 14-1). La justificación ética de estas exenciones es que la investigación conlleva poco riesgo, casi todas las personas consentirían esa investigación y la obtención del consentimiento de cada participante haría que fueran estudios prohibitivamente caros o difíciles. Sin embargo, muchos CEIC exigen que los investigadores envíen algunos datos sobre el proyecto para verificar que se cumplen los criterios de la exención.
- Un CEIC puede permitir que algunas investigaciones de riesgo mínimo sean sometidas a una revisión rápida por parte de un solo revisor, en lugar de por el comité entero (tabla 14-2). La página web de la Office for Human Research Protections enumera los tipos de investigación idóneos para una revisión rápida (6). El concepto de mínimo riesgo para los participantes tiene un papel clave en las leyes federales, como se indica en la tabla 14-2 . Se define el mínimo riesgo como el que «normalmente se encuentra en la vida diaria o durante la realización de pruebas físicas o psicológicas habituales». Deben tenerse en cuenta tanto la magnitud como la probabilidad del riesgo. El CEIC debe juzgar si puede considerarse que un proyecto específico conlleva un mínimo riesgo.

---

1. Estudios, entrevistas u observaciones de comportamiento público, salvo que:
    - Se pueda identificar a los participantes *y*
    - La divulgación de las respuestas de los participantes pueda suponerles un riesgo de responsabilidad legal o dañar su reputación, posición económica o empleo. Por ejemplo, no están exentos los cuestionarios sobre aspectos como la adicción a drogas, la depresión, las conductas de riesgo del VIH y la inmigración ilegal.
2. Estudios de registros, datos o muestras ya existentes, siempre que:
    - Estén disponibles públicamente (p. ej., conjuntos de datos liberados por agencias estatales y federales) O
    - Registrados por el investigador de tal manera que no se puede identificar a los participantes, por ejemplo, porque el investigador no puede obtener la clave del código.
3. Investigación sobre prácticas formativas normales.

---

1. Algunos procedimientos de investigación que conllevan un riesgo mínimo, como:
    - Recopilación de muestras mediante punción venosa, recogida de saliva o esputo, o muestras de piel o de mucosas obtenidas con hisopo.
    - Recopilación de muestras mediante procedimientos incruentos empleados habitualmente en la práctica clínica, como electrocardiogramas y resonancia magnética. Las radiografías, que exponen a los participantes a la radiación, precisan la

revisión completa por el CEIC.

- Investigación que conlleva datos, registros o muestras que han sido recopilados o que lo serán con fines clínicos.
- Investigaciones que usan estudios o entrevistas que no están exentos de la revisión por el CEIC.

2. Modificaciones leves de protocolos de investigaciones previamente aprobadas.
3. Renovación de la autorización por el CEIC para estudios que ya hayan finalizado, excepto el análisis de los datos o el seguimiento a largo plazo.

Consentimiento informado y voluntario

Los investigadores deben obtener el consentimiento informado y voluntario de los participantes en la investigación.

Comunicación de información a los participantes

Las leyes federales precisan que los investigadores comenten con los posibles participantes diversos temas, como:

- Naturaleza del proyecto de investigación. Al posible participante debe decírsele explícitamente que se está realizando una investigación, cuál es su objetivo y a quién se va a incluir como participante. No es necesario que se describa la hipótesis específica del estudio.
- Procedimientos del estudio. Los participantes deben saber qué se les va a pedir que hagan en el proyecto de investigación. En la práctica, se les debe comunicar cuánto tiempo se les pedirá y con cuánta frecuencia. Deben identificarse como tales todos los procedimientos que no formen parte de la asistencia clínica habitual. Si el estudio conlleva enmascaramiento o aleatorización, deben explicarse estos conceptos con palabras que pueda entender el participante. En la investigación con entrevistas o cuestionarios, debe informarse al participante de los temas que se van a tratar.
- Riesgos y posibles beneficios del estudio y alternativas a participar en el estudio. Deben describirse con palabras no técnicas los perjuicios y beneficios médicos, psicosociales y económicos. También debe informarse a los posibles participantes de las alternativas a la participación, por ejemplo, si pueden participar en un ensayo clínico fuera del estudio. Ha salido a la luz la preocupación de que a menudo la información que se proporciona a los participantes subestima los perjuicios y exagera los beneficios (7). Por ejemplo, a veces se describe que la investigación sobre un nuevo fármaco ofrece beneficios a los participantes. Sin embargo, las nuevas intervenciones más prometedoras, a pesar de unos resultados preliminares alentadores, no tienen ventajas significativas sobre el tratamiento habitual. Los participantes habitualmente tienen un «concepto erróneo terapéutico» de que la intervención está diseñada para proporcionarles un beneficio personal (8). Los investigadores deben dejar claro que no se sabe si el fármaco del estudio es más eficaz que el tratamiento habitual, y que fármacos prometedores pueden causar graves daños.

Formularios de consentimiento

Generalmente hacen falta formularios de consentimiento por escrito para documentar que se ha producido el proceso del consentimiento informado: conversaciones entre el investigador y el participante. Estos formularios deben contener la información necesaria que se ha señalado en la sección previa. Por otro lado, puede usarse un formulario abreviado que establece que los elementos necesarios del consentimiento informado se han presentado de forma oral. Si se usa el formulario abreviado, en la presentación oral debe haber un testigo, que debe firmar el formulario de consentimiento abreviado junto con el participante.

Los CEIC suelen tener modelos de formularios de consentimiento que prefieren que usen los investigadores. Los CEIC pueden solicitar que se divulgue más información de la que se solicita en las leyes federales.

Comprensión por parte de los participantes de la información desvelada

Los participantes en una investigación habitualmente tienen graves errores de concepto sobre los objetivos de la investigación y los procedimientos y riesgos del protocolo específico (9). En conversaciones y formularios de consentimiento, los investigadores deben evitar el uso de lenguaje técnico y de frases complicadas. Se ha criticado a los CEIC por centrarse en exceso en los formularios de consentimiento, en lugar de en si los participantes han entendido la información pertinente (9). Entre las estrategias para aumentar la comprensión de los participantes se encuentran contar con un miembro del equipo del estudio o un formador neutral que pase más tiempo hablando uno por uno con los participantes del estudio, simplificar los formularios de consentimiento, usar un formato de pregunta y

respuesta, proporcionar la información en varias visitas, y usar cintas de audio o vídeo (10). En la investigación que conlleva un riesgo importante o que es controvertida, los investigadores deben considerar la evaluación de la comprensión por los participantes y la documentación de que el participante puede responder correctamente preguntas sobre los aspectos fundamentales de la investigación (11 , 12).

Naturaleza voluntaria del consentimiento

El consentimiento éticamente válido debe ser voluntario e informado. Los investigadores deben reducir al mínimo la posibilidad de coacción o de influencia indebida. Son ejemplos de influencia indebida los pagos excesivos a los participantes y la inclusión de estudiantes como participantes en la investigación. La influencia indebida plantea problemas éticos si lleva a los participantes a rebajar significativamente los riesgos de un proyecto de investigación, o si socava gravemente su capacidad de negarse a participar. Los participantes deben entender que rechazar la participación en el estudio no afectará a su asistencia médica y que pueden retirarse del proyecto en cualquier momento.

Excepciones al consentimiento informado

Algunos estudios científicamente importantes serían difíciles o imposibles de realizar si se debiera obtener el consentimiento informado de todos los participantes.

CASO 14-2. Investigación con muestras de sangre de recién nacidos

Poco después del nacimiento se realiza a los lactantes una punción en el talón para obtener sangre en un papel de filtro para realizar cribado de enfermedades genéticas. En la mayoría de los estados no es necesario el permiso de los padres para este cribado obligatorio; por lo tanto, las muestras representan a toda la población de recién nacidos. Las muestras que sobran después del cribado clínico han sido útiles para investigar causas genéticas de malformaciones neonatales y parto pretérmino, exposición a factores ambientales durante el embarazo, e interacciones entre genes y entorno.

Investigación con muestras y datos sobrantes sin identificación

No hace falta el consentimiento informado y la revisión por el CEIC para utilizar muestras sin identificación en la investigación (v. tabla 14-1), aunque muchos CEIC siguen exigiendo que los investigadores les notifiquen dichas investigaciones. Cuando se envía una investigación original para su publicación, muchas revistas exigen que los autores declaren que un CEIC autorizó el protocolo o determinó que no era necesaria la revisión.

Exención del consentimiento informado

Algunos proyectos de investigación valiosos necesitan información y muestras previas con identificación. Estos estudios no cumplen los criterios para la exención de la revisión por el CEIC, aunque pueden cumplir los criterios para una exención del consentimiento informado.

CASO 14-2. Investigación con muestras de sangre de recién nacidos *(cont.)*

Un equipo de investigación querría utilizar muestras de sangre de recién nacidos con identificación para estudiar la asociación entre la exposición ambiental de la madre a determinados productos químicos y el bajo peso al nacimiento, la prematuridad y la muerte perinatal. Los investigadores pueden relacionar las muestras identificadas con certificados de nacimiento, certificados de defunción e informes hospitalarios. Debido al gran número de niños a los que se debe estudiar para conseguir una potencia adecuada para detectar asociaciones, no sería viable obtener la autorización de los padres o tutores.

De acuerdo con las leyes federales, los CEIC pueden conceder exenciones de consentimiento informado si se aplican todas las condiciones de la tabla 14-3 . La mayoría de los CEIC eximiría del consentimiento al estudio propuesto de exposiciones ambientales maternas.

Justificación de las excepciones al consentimiento informado

Algunos proyectos de investigación con importancia científica conllevan riesgos tan bajos que el consentimiento sería laborioso, a la vez que sería poco útil para proteger a los participantes en la investigación. Todos los pacientes se han beneficiado del conocimiento obtenido de estudios de investigación en los que se han utilizado historias y muestras ya existentes. La justicia en el sentido de reciprocidad indica que las personas que reciben dichos beneficios estarían deseando participar en una investigación similar de riesgo muy bajo para beneficiar a otras personas.

Objeciones a las excepciones al consentimiento informado

Aunque las leyes federales permiten la utilización de muestras de sangre de recién nacidos sin identificación para la investigación sin permiso de los padres, hay una oposición significativa del público.

CASO 14-2. Investigación con muestras de sangre de recién nacidos *(cont.)*

Padres de varios estados han objetado el almacenamiento de muestras para una investigación no especificada sin su permiso o sin la oportunidad de retirarse de la investigación, lo que ha llevado a pleitos en dos estados. Los demandantes no se opusieron a la recogida de sangre para el cribado neonatal, aunque sí objetaron que incluso la eliminación de la identificación de las muestras no abordaba sus preocupaciones sobre la pérdida de la privacidad y la autonomía.

Como estas objeciones a la investigación podrían socavar la aceptación clínica del cribado neonatal, los estados están ofreciendo cada vez más a los padres la oportunidad de rechazar los usos de investigación de las muestras neonatales obtenidas en los programas de cribado estatales. Esta atención a los deseos de los padres puede ir más allá de lo que precisan las leyes federales sobre la investigación. Así, lo que está permitido legalmente en investigación podría no siempre ser aceptable desde el punto de vista ético, particularmente en investigaciones sobre temas delicados.

---

1. La investigación solo supone un riesgo mínimo para los participantes; *y*
2. La exención o alteración no afectará negativamente a los derechos y el bienestar de los participantes; *y*
3. La investigación no se podría llevar a cabo de manera práctica sin la exención; *y*
4. Cuando proceda, se deberá entregar a los participantes toda la información adicional que corresponda después de su participación. Esta provisión permite la realización de investigaciones que suponen cierto grado de engaño, por ejemplo, cuando la comunicación de la finalidad de la investigación socavaría la validez del estudio.

Participantes que carecen de capacidad de tomar decisiones

Cuando los participantes no son capaces de dar su consentimiento informado, la autorización para participar en el estudio debe obtenerse del representante legalmente autorizado del participante (un progenitor o el tutor en el caso de niños pequeños). Además, el protocolo debe estar sujeto a un examen adicional para asegurar que la pregunta de la investigación no puede estudiarse en una población que sí es capaz de dar ese consentimiento.

Minimización de los riesgos

Los investigadores deben prever los riesgos que se podrían producir en los proyectos de investigación y reducirlos, por ejemplo identificando y excluyendo a las personas que son muy susceptibles a los acontecimientos adversos, con un seguimiento adecuado para detectar acontecimientos adversos, y utilizando como alternativa mediciones menos invasivas. Un aspecto importante de la reducción al mínimo de los riesgos es mantener la confidencialidad de los participantes.

Confidencialidad

La violación de la confidencialidad puede causar estigmas o discriminación, sobre todo si la investigación aborda asuntos sensibles, como las actitudes y las prácticas sexuales, el consumo de alcohol o drogas, la conducta ilegal y las enfermedades psiquiátricas. Las estrategias para proteger la confidencialidad comprenden la codificación de los datos de la investigación, la protección o destrucción de la clave que identifica a los participantes, y la limitación del personal que tiene acceso a los identificadores. Sin embargo, los investigadores no deben realizar promesas ilimitadas de confidencialidad. Esta puede quedar invalidada si se realiza una auditoría de los registros de la investigación, o si se identifican enfermedades que se deben notificar de acuerdo con la legislación, como maltrato infantil, determinadas enfermedades infecciosas y amenazas graves de violencia. En los proyectos en los que se puede prever dicha notificación, el protocolo debe especificar cómo debe responder el personal que realiza el estudio, y se debe informar a los participantes de estos planes.

Los investigadores pueden evitar citaciones en disputas legales obteniendo certificados de confidencialidad del Servicio de Salud Pública (Public Health Service) (13) que les permitan retener datos de investigación identificables si se enfrentan a una citación o una orden judicial para desvelarlos. Sin embargo, estos certificados no se han estudiado de manera generalizada en los fallos judiciales, no se aplican a las auditorías por agencias financiadoras o la FDA, y no impiden que el investigador notifique voluntariamente información

relacionada con maltrato de niños o ancianos, violencia doméstica o enfermedades infecciosas de notificación obligatoria. No es necesario que la investigación esté financiada por el gobierno federal para que reciba un certificado de confidencialidad.

Ley de privacidad sanitaria (HIPAA)

La ley federal *Health Privacy Rule* (ley de privacidad sanitaria, conocida habitualmente como HIPAA, acrónimo de *Health Insurance Portability and Accountability Ac*t [ley de portabilidad y responsabilidad de seguros sanitarios]) protege la información sanitaria identificable a nivel individual, a la que se denomina información sanitaria protegida. De acuerdo con la ley de privacidad, los participantes deben firmar una autorización para el uso de información sanitaria protegida en un proyecto de investigación (14). Este formulario de autorización de la HIPAA se suma al formulario de consentimiento informado que exige el CEIC. Los investigadores deben obtener la autorización para cada uso de información protegida para la investigación; no se permite el consentimiento general para la investigación futura. No hace falta la autorización si los datos no se pueden identificar y en algunas otras situaciones determinadas. Los investigadores deben ponerse en contacto con el CEIC si tienen dudas sobre la ley de privacidad y en qué aspectos difiere de las leyes federales sobre la protección de sujetos humanos.

Volver al principio

▪ PARTICIPANTES DE LA INVESTIGACIÓN QUE PRECISAN OTRAS PROTECCIONES ADICIONALES

Algunos participantes podrían tener «mayor riesgo de ser utilizados en la investigación de un modo éticamente inadecuado», por dificultad en dar su consentimiento voluntario e informado o por mayor tendencia a sufrir acontecimientos adversos (15).

Tipos de vulnerabilidad

La identificación de los diferentes tipos de vulnerabilidad permite que los investigadores adopten medidas de seguridad adaptadas al tipo específico de vulnerabilidad.

Alteraciones cognitivas o de la comunicación

Las personas con deterioro de la cognición o la comunicación podían tener dificultad para comprender la información sobre un estudio y para sopesar los riesgos y beneficios del estudio.

Diferencias de poder

Las personas ingresadas en centros o instituciones, como los presos y los ancianos que viven en residencias, podrían sentirse presionados a participar en la investigación y a delegar la autorización en las personas que controlan su rutina diaria. Estas personas vulnerables podrían no darse cuenta de que pueden negarse a participar en el estudio sin represalias por la autoridad ni peligro para otros aspectos de sus vidas cotidianas.

Si el investigador del proyecto de investigación es también un médico que trata a los participantes, estos podrían encontrar difícil declinar la participación en el estudio por miedo a que el médico estuviera menos interesado en su asistencia. De manera similar, los estudiantes y los médicos en formación podrían sentirse presionados a participar en una investigación realizada por sus instructores o superiores.

Inconvenientes sociales y económicos

Las personas con un nivel socioeconómico bajo o con un acceso escaso a la asistencia sanitaria podrían participar en un estudio de investigación para obtener un pago o asistencia médica, incluso en el caso de que consideraran que los riesgos son inaceptables si tuvieran una mejor situación económica. Los participantes con poca formación pueden no comprender la información sobre el estudio, o pueden ser más susceptibles a la influencia por otras personas.

Protecciones para los participantes vulnerables

Se pueden encontrar las leyes federales sobre la investigación con participantes vulnerables en la página web de la oficina federal para la protección de la investigación en seres humanos (Office for Human Research Protections) (3).

Investigación en niños

Los investigadores deben obtener la autorización de los padres y la aprobación del niño, si su desarrollo es adecuado. La investigación con niños que conlleve más del mínimo riesgo solo es permisible:

- Si ofrece la perspectiva de un beneficio directo al niño O
- Si el aumento respecto al riesgo mínimo es pequeño y es probable que la investigación permita obtener «conocimiento generalizable de importancia vital sobre el trastorno o la enfermedad del niño».

Investigación en presos

Los presos podrían no sentirse con libertad para rechazar su participación en una investigación, y se podría ejercer sobre ellos una influencia indebida con pagos en efectivo, modificaciones de la rutina de la prisión o consideraciones de puesta en libertad condicional. Las leyes federales limitan los tipos de investigación permitidos en prisioneros y precisan una revisión más estricta por el CEIC y la aprobación por el Department of Health and Human Services (Ministerio de sanidad y servicios humanos).

Investigación en mujeres embarazadas, fetos y embriones

La investigación que no ofrece ninguna perspectiva de beneficio directo para el feto se permite solo si «la finalidad de la investigación es el desarrollo de un conocimiento biomédico importante que no se puede obtener por ningún otro medio». La investigación que ofrece la perspectiva de un beneficio directo solo para el feto precisa el consentimiento informado tanto del padre como de la embarazada, aunque la investigación que ofrece a los niños la perspectiva de un efecto beneficioso directo precisa la autorización de tan solo un progenitor. Se han criticado estas restricciones, porque disuaden de una investigación que reforzaría la base de datos científicos para la asistencia clínica de las embarazadas y sus fetos.

Volver al principio

▪ RESPONSABILIDADES DE LOS INVESTIGADORES

Actualmente siguen produciéndose denuncias por conducta gravemente impropia en la investigación.

CASO 14-3. Efectos adversos cardíacos del rofecoxib

En 2002 se publicaron los resultados del ensayo aleatorizado y controlado VIGOR. En este estudio se comparó un nuevo antiinflamatorio no esteroideo inhibidor selectivo de la COX-2, el rofecoxib, con un fármaco no selectivo y más antiguo, el naproxeno (16). El fabricante del rofecoxib patrocinó el estudio. El rofecoxib produjo significativamente menos complicaciones gastrointestinales que el naproxeno (2,1 y 4,5 por cada 1 000 pacientes/año, respectivamente), a la vez que tenía una eficacia similar en el dolor de la artrosis. En el grupo de rofecoxib se observaron también más infartos de miocardio (el 0,4 %, en comparación con el 0,1 %). Después de esta publicación se prescribió ampliamente el rofecoxib, con ventas anuales mayores de 2 500 millones de dólares. Antes de la publicación del artículo se habían notificado a la FDA otros tres infartos de miocardio en el grupo de rofecoxib, aunque no se notificaron a los autores del artículo, que trabajaban en universidades, ni a la revista. Dos autores que eran empleados del fabricante conocían estos casos adicionales. La revista en la que se publicó el estudio VIGOR publicó posteriormente una expresión de su preocupación de que el «artículo no hubiera representado con exactitud los datos de seguridad de que se disponía cuando se estaba revisando el artículo para su publicación» (17). Además de retener datos desfavorables, en la publicación se estableció una fecha de corte anterior para los acontecimientos adversos cardiovasculares que para los acontecimientos adversos gastrointestinales, sin desvelar esto a la revista ni a los autores académicos del estudio, lo que sesgó los resultados a favor del rofecoxib.

Posteriormente, en otro estudio aleatorizado se demostró que el rofecoxib producía significativamente más infartos de miocardio y accidentes cerebrovasculares que el naproxeno (18), y el fabricante retiró voluntariamente el fármaco del mercado.

En otras publicaciones muy influyentes, los investigadores inventaron o alteraron intencionadamente los datos, por ejemplo, cuando propusieron un vínculo entre la vacuna triple vírica y el autismo infantil, y cuando afirmaron que habían obtenido una línea de células progenitoras humanas utilizando trasplante nuclear en células somáticas (19 , 20). Esta conducta impropia socava la confianza del público y de los médicos en la investigación, y amenaza la financiación pública de la investigación.

Conducta científica impropia

La oficina federal para la integridad de la investigación (Federal Office for Research Integrity) define la conducta impropia en la investigación como la invención, la falsificación y el plagio (21).

- La invención es crear resultados y registrarlos o notificarlos.
- La falsificación es la manipulación de materiales de la investigación, el equipo o los procedimientos, o la modificación o la omisión

de datos o resultados, de manera que el informe de la investigación represente erróneamente los hallazgos reales.

- El plagio es la apropiación de las ideas, resultados o palabras de otra persona sin reconocer correctamente los méritos.

En esta definición federal, la conducta impropia debe ser intencionada, en el sentido de que los perpetradores son conscientes de que su conducta no es adecuada. En el caso 14-3 no se pudo demostrar la falsificación intencionada de los hallazgos. La conducta inadecuada excluye el error sincero y las legítimas diferencias de opinión científicas, que forman parte normal del proceso de investigación. La definición federal no aborda otras acciones incorrectas, como la doble publicación, la negativa a compartir los materiales de la investigación y el acoso sexual; los centros de investigación deben afrontarlos aplicando otras políticas.

Cuando se denuncia una conducta impropia en la investigación, tanto el organismo de financiación federal como el centro del investigador tienen la responsabilidad de realizar una investigación clara y puntual (22). Durante una investigación, tanto los delatores como los científicos acusados tienen derechos que hay que respetar. Los primeros necesitan protección frente a represalias, y a los segundos se les debe informar de los cargos y se les debe dar la oportunidad de responder. El castigo por una conducta impropia científica probada puede conllevar la suspensión de una subvención, la exclusión para subvenciones futuras y otras sanciones administrativas, académicas, penales o civiles.

CASO 14-3. Efectos adversos cardíacos del rofecoxib *(cont.)*

Muchos pacientes que habían tomado rofecoxib y sufrieron un infarto de miocardio demandaron al fabricante. Durante el proceso legal se presentaron correos electrónicos internos del promotor en los que se indicaba que los borradores de muchos artículos sobre rofecoxib fueron realizados por empleados o consultores de la compañía, y muchas veces se invitaba a investigadores académicos a que fueran los primeros autores solo después de haber elaborado el borrador del manuscrito. Con frecuencia no se citaba y reconocía como autores a los empleados que habían elaborado los borradores de los artículos.

Autoría

Para cumplir los criterios de autoría, los investigadores deben hacer contribuciones importantes a:

- La concepción y el diseño del estudio, o el análisis y la interpretación de los datos, *y*
- La elaboración del borrador o la revisión del artículo, *y*
- La autorización final del manuscrito (23).

La autoría invitada y la autoría «fantasma» no son éticas. Se reconoce como autores a los autores invitados u honoríficos a pesar de haber hecho contribuciones tan solo triviales al artículo, por ejemplo, proporcionando el reconocimiento que conlleva su nombre, el acceso a los participantes, los reactivos, la ayuda de laboratorio o la financiación. En el caso 14-3 no es adecuado que las personas lleguen a ser autores después de que haya finalizado el estudio, de que se hayan analizado los datos y de que se haya elaborado el primer borrador. Los autores «fantasma» hacen contribuciones importantes a un artículo, pero no se los enumera como autores. Generalmente son empleados de las compañías farmacéuticas o de compañías de elaboración de textos médicos. La omisión de los escritores «fantasma» orienta erróneamente a los lectores para que infraestimen la participación de la compañía en el manuscrito. De acuerdo con un estudio, el 25 % de los artículos de investigación originales en revistas médicas generales de elevado índice de impacto tienen autores invitados, y el 12 % tienen autores «fantasma» (24).

Frecuentemente surgen desacuerdos sobre quién debe ser el autor o el orden de los autores. Estos temas se comentan y deciden mejor al principio de un proyecto. Los cambios en cuanto a la autoría deben negociarse si se ha tomado la decisión de cambiar las responsabilidades del trabajo. Se han propuesto sugerencias para llevar a cabo esas negociaciones diplomáticamente (25). Como no hay acuerdo sobre los criterios de la posición de los autores, en algunas revistas se describen en el artículo publicado las contribuciones al proyecto de cada uno de los autores.

Conflictos de intereses

Los intereses principales del investigador deben ser ofrecer respuestas válidas a preguntas científicas importantes y proteger la seguridad de los participantes. Los investigadores podrían tener otros intereses, como su reputación o sus ingresos, que entran en contradicción con los objetivos principales de la investigación y podrían deteriorar su objetividad o socavar la confianza del público en la investigación (26).

Tipos de conflictos de intereses

- Conflictos de intereses económicos. Los estudios de fármacos, dispositivos y pruebas nuevos los suele financiar la industria. La

cuestión ética es que determinados vínculos económicos podrían producir sesgo en el diseño y la realización del estudio, la interpretación excesiva de los resultados positivos o la falta de publicación de resultados negativos (27 , 28). Si los investigadores son titulares de patentes de la intervención en estudio o tienen opciones de participaciones en la compañía que fabrica un fármaco o dispositivo en estudio, podrían conseguir grandes beneficios económicos si se demuestra que el tratamiento es eficaz, además de su compensación por la realización del estudio. Finalmente, la recepción de grandes honorarios por consultoría, emolumentos o regalos en especie podría sesgar el juicio de un investigador a favor del producto de la compañía.

- Funciones dobles de médicos-investigadores. Si un investigador es el médico personal de un participante elegible para la investigación, podrían entrar en conflicto las funciones de médico e investigador. Los pacientes podrían temer que su asistencia futura peligrara si se negaran a participar en la investigación, y podrían no distinguir entre investigación y tratamiento. Además, lo que es mejor para un participante concreto podría diferir de lo que es mejor para el proyecto de la investigación.

Responder a los conflictos de intereses

Se deben desvelar todos los conflictos de intereses, y algunos tienen tanta capacidad de introducir sesgos en los resultados de la investigación que se deben corregir o evitar.

- Reducir la probabilidad de sesgo. En ensayos clínicos bien diseñados, varias precauciones habituales ayudan a mantener controlados intereses contradictorios. Se puede ocultar a los investigadores la intervención que recibe un participante para evitar el sesgo en la valoración de las respuestas. Un comité de supervisión de los datos y la seguridad (CSDS) independiente (cap. 11), cuyos miembros no tengan conflictos de intereses, puede revisar los datos provisionales y poner fin al estudio si los datos proporcionan pruebas convincentes de beneficio o perjuicio. El proceso de revisión por especialistas de subvenciones, resúmenes y manuscritos también contribuye a reducir el sesgo.

- Separar las funciones que entran en conflicto. Los médicos deben separar la función de investigador en un proyecto de la función de médico que presta la asistencia médica al participante. En general, los médicos no deben incluir a sus propios pacientes en un estudio de investigación del que son coinvestigadores. Si se incluye a estos pacientes, un miembro del equipo de investigación que no sea el médico que trata a los pacientes debe encargarse de las discusiones para el consentimiento.

- Control del análisis y las publicaciones. En investigaciones financiadas por una compañía farmacéutica, los investigadores deben asegurarse de que el contrato les da control sobre los datos primarios y el análisis estadístico, así como la libertad para publicar los hallazgos, tanto si el fármaco en investigación es eficaz como si no lo es (27 , 28). El investigador tiene una obligación ética de responsabilizarse de todos los aspectos de la investigación. El patrocinador puede revisar los manuscritos, hacer sugerencias y asegurar que las solicitudes de patente se han enviado antes de que se remita el artículo a una revista. Sin embargo, el patrocinador no debe tener poder de veto ni censura de la publicación, ni insistir en un enunciado específico en el manuscrito.

- Revelar conflictos de intereses. Los centros de investigación exigen que se comuniquen los conflictos de intereses a una oficina encargada de este asunto. Los NIH y otras agencias que financian estudios, los CEIC locales, las reuniones científicas y las revistas médicas exigen la declaración de los conflictos de intereses cuando se solicitan subvenciones o se envíen resúmenes y artículos. Aunque la comunicación por sí sola es, a menudo, una respuesta inadecuada a conflictos de intereses graves, podría desanimar a los investigadores a que realizaran prácticas problemáticas desde el punto de vista ético, y permite que los revisores y los lectores de los artículos de revistas evalúen la posibilidad de una influencia indebida.

- Manejo del conflicto de intereses. Si un estudio concreto presenta conflictos de intereses significativos, el centro de investigación, la agencia financiadora o el CEIC pueden exigir salvaguardas adicionales, como un seguimiento más estrecho del proceso de consentimiento informado o una modificación de la función del investigador que ha entrado en conflicto.

- Prohibir determinadas situaciones. Para reducir al mínimo los conflictos de intereses, los financiadores y los centros académicos podrían prohibir que el titular de la patente de una intervención o un empleado de la compañía que fabrica la intervención actuara como investigador principal en un estudio clínico.

Volver al principio

▪ CUESTIONES ÉTICAS ESPECÍFICAS PARA DETERMINADOS TIPOS DE INVESTIGACIÓN

Ensayos clínicos aleatorizados

Aunque los ensayos clínicos aleatorizados constituyen el diseño más riguroso para evaluar intervenciones (cap. 10), presentan problemas éticos especiales por dos motivos: la intervención está determinada por el azar y, al contrario de los estudios de observación,

los investigadores realizan una intervención en los participantes. Una justificación ética para asignar el tratamiento mediante aleatorización es que las intervenciones del estudio están en equilibrio, un concepto que intuitivamente parece claro, pero sobre el que hay un intenso debate y que es imposible de definir de manera precisa (29). Debe haber una incertidumbre o controversia genuina sobre qué grupo del estudio es superior, de manera que los participantes no sufran un perjuicio significativo si se permite que su asistencia esté determinada por la aleatorización y no por su médico personal. El equilibrio no exige que haya un balance exacto entre los grupos del estudio.

Los participantes en un estudio clínico reciben una intervención cuyos efectos adversos se podrían desconocer. Por lo tanto, en estos ensayos hace falta un seguimiento cuidadoso para asegurarse de que los participantes no sufran un perjuicio excesivo. El investigador tiene la responsabilidad de establecer métodos cuidadosos para la evaluación de los acontecimientos adversos (caps. 10 y 11). En la mayoría de los estudios, esto incluye el establecimiento de un comité de supervisión de los datos y la seguridad independiente que revise de manera intermitente los datos del estudio y que pueda poner fin al ensayo si hay un perjuicio inesperado asociado a la intervención (cap. 11).

Las intervenciones en los grupos testigo también hacen surgir cuestiones éticas. Si hay un tratamiento estándar o eficaz para una enfermedad, el grupo testigo debe recibirlo (cap. 11). Sin embargo, puede justificarse el tratamiento con un placebo de los testigos en ensayos a corto plazo que no supongan riesgos importantes para los participantes, como estudios de hipertensión leve y de dolor leve de resolución espontánea. Debe informarse a los participantes de las intervenciones eficaces de que se dispone fuera del estudio.

No es ético continuar un ensayo clínico si existen pruebas de que la intervención de uno de los grupos es más segura o más eficaz. Además, sería un error continuar un ensayo que no responderá a la pregunta de la investigación porque hay pocos participantes, pocas respuestas o una elevada frecuencia de abandono. El análisis periódico de los datos provisionales en un ensayo clínico por parte de un comité de supervisión de los datos y de la seguridad independiente puede determinar si un ensayo debería finalizarse prematuramente por estos motivos (30). Estos análisis provisionales no deben realizarlos los investigadores, porque, si no desconocen los hallazgos provisionales, puede producirse sesgo si el estudio continúa. Antes de incluir a los participantes deben especificarse los procedimientos para examinar los datos provisionales (cap. 11).

Los ensayos clínicos que se realizan en países en desarrollo plantean otros dilemas éticos (cap. 18).

Investigación con datos y muestras anteriormente recopilados

La investigación con datos obtenidos previamente y con muestras almacenadas ofrece la posibilidad de hacer descubrimientos significativos. Por ejemplo, el estudio del ADN en un gran número de muestras biológicas almacenadas que están vinculadas con datos clínicos podría identificar genes que aumentan la probabilidad de desarrollar una enfermedad, de tener mal pronóstico o de responder a un tratamiento concreto. Los grandes bancos biológicos de muestras de sangre y de tejidos permiten realizar estudios futuros sin tener que recopilar más muestras. La investigación con datos y muestras recopiladas anteriormente no plantea riesgos físicos a los participantes. Sin embargo, sí existen cuestiones éticas. El consentimiento para estudios futuros no especificados es problemático, porque nadie puede prever qué tipo de investigación podría realizarse más adelante. Además, los participantes podrían poner objeciones al uso futuro de los datos y muestras en determinadas formas. Si se producen violaciones de la confidencialidad, podrían llevar a estigma y discriminación. Los grupos que participen en la investigación podrían resultar perjudicados incluso si no lo fueran los participantes individuales.

Cuando se recogen muestras biológicas, los formularios de consentimiento deben permitir que los participantes acepten o rechacen algunos tipos de investigación futura con las muestras. Por ejemplo, los participantes podrían estar de acuerdo en permitir el uso de sus muestras:

- Para investigaciones futuras que hayan sido autorizadas por un CEIC y un panel de revisión científica, o
- Solo para una investigación en condiciones específicas, o
- Solo en el estudio de investigación actual, no en estudios futuros.

Los participantes también deben saber si se compartirán con otros investigadores datos y muestras que se puedan identificar. Además, los participantes deben saber que los descubrimientos de la investigación obtenidos con sus muestras podrían patentarse y transformarse en productos comerciales.

Volver al principio

▪ OTROS ASPECTOS

Pago a los participantes en una investigación

Los participantes en una investigación clínica merecen un pago por su tiempo y su esfuerzo, y el reembolso de gastos extra como el transporte o el cuidado de los niños. En la práctica también podría ser necesaria una compensación para reunir y conservar a los participantes. Una práctica habitual es ofrecer mayor compensación en los estudios que conllevan muchas incomodidades o riesgos. Sin embargo, los incentivos plantean también cuestiones éticas sobre una posible inducción indebida. Si se paga más a los participantes por entrar en investigaciones de riesgo, las personas de menor nivel socioeconómico podrían verse inducidas a despreciar estos riesgos. Para evitar la influencia indebida, se ha sugerido que solo se compense a los participantes por los gastos y el tiempo reales, como una tarifa horaria correspondiente a un trabajo no especializado (31).

Volver al principio

▪ RESUMEN

- Los investigadores deben garantizar que sus proyectos cumplan los principios éticos de respeto a las personas, beneficencia y justicia.
- Los investigadores deben asegurarse de que la investigación cumpla con los requisitos de las leyes federales aplicables, de los cuales las principales características son el consentimiento informado de los participantes y la revisión por un CEIC. Durante el proceso del consentimiento informado, los investigadores deben explicar a los posibles participantes la naturaleza del proyecto y los riesgos, posibles beneficios y alternativas. Los investigadores deben garantizar la confidencialidad de la información de los pacientes, en cumplimiento de la ley de privacidad de datos sanitarios (HIPAA).
- Las poblaciones vulnerables, como niños, presos, mujeres embarazadas y personas con defectos cognitivos o problemas sociales, necesitan una protección adicional.
- Los investigadores deben tener integridad ética. No deben cometer una conducta impropia científica, que, según las regulaciones, se define como invención, falsificación o plagio. Los investigadores deben comunicar y manejar adecuadamente los conflictos de intereses y deben seguir criterios para una autoría correcta, citándose a sí mismos como autores de un manuscrito solo si han hecho contribuciones intelectuales importantes, y garantizando que se cite como autores a todas las personas que hayan hecho contribuciones importantes a un manuscrito.
- En algunos tipos de investigación deben abordarse otras cuestiones éticas. En los ensayos clínicos aleatorizados, las ramas de la intervención deben estar equilibradas, los grupos testigo deben recibir intervenciones adecuadas y no debe continuarse el ensayo una vez que se demuestre que un grupo es más eficaz o perjudicial. Cuando la investigación se realiza con datos o muestras recopilados anteriormente, se debe prestar una especial atención a la confidencialidad.

Volver al principio

BIBLIOGRAFÍA

1. Jones JH. The Tuskegee syphilis experiment. In: Emanuel EJ, Grady C, Crouch RA, et al., editors. Oxford textbook of research ethics. New York: Oxford University Press, 2008, 86-96. Citado aquí

2. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: Ethical principles and guidelines for the protection of human subjects of biomedical and behavioral research. 1979. Available at: www.hhs.gov/ohrp/humansubjects/guidance/belmont.html, accessed 8/27/12. Citado aquí

3. Department of Health and Human Services. Protection of human subjects 45 CFR part 46. 2005. Available at: www.dhhs.gov/ohrp/humansubjects/guidance/45cfr46.html, accessed 9/27/12. Citado aquí

4. Emanuel EJ, Menikoff J. Reforming the regulations governing research with human subjects. N Engl J Med 2011; 365:1145-50. Citado aquí

5. Lo B, Barnes M. Protecting research participants while reducing regulatory burdens. JAMA 2011;306:2260-2261. Citado aquí

6. Department of Health and Human Services. Protocol review. 2005. Available at: www.dhhs.gov/ohrp/policy/protocol/index.html, accessed 9/27/12. Citado aquí

7. King NMP, Churchill LR. Assessing and comparing potential benefits and risks of harm. In: Emanuel EJ, Grady C, Crouch RA, et al., editors. The Oxford textbook of clinical research ethics. New York: Oxford University Press, 2008, 514-526. Citado aquí

8. Henderson GE, Churchill LR, Davis AM, et al. Clinical trials and medical care: defining the therapeutic misconception. PLoS Med 2007;4:e324. Citado aquí

9. Federman DD, Hanna KE, Rodriguez LL. Responsible research: a systems approach to protecting research participants. 2002. Available at: www.nap.edu/catalog.php?record_id=10508, accessed 9/29/12. Citado aquí

10. Flory J, Emanuel E. Interventions to improve research participants' understanding in informed consent for research: a systematic review. JAMA 2004;292:1593-1601. Citado aquí

11. Lomax GP, Hall ZW, Lo B. Responsible oversight of human stem cell research: the California Institute for Regenerative Medicine's medical and ethical standards. PLoS Med 2007;4:e114. Citado aquí

12. Woodsong C, Karim QA. A model designed to enhance informed consent: experiences from the HIV prevention trials network. Am J Public Health 2005;95:412-419. Citado aquí

13. Wolf LE, Dame LA, Patel MJ, et al. Certificates of confidentiality: legal counsels' experiences with perspectives on legal demands for reseasch data. J Empir Res Hum Res Ethics 2012;7:1-9. Citado aquí

14. Nass SJ, Leavitt LA, Gostin LO. Beyond the HIPAA Privacy Rule: enhancing privacy, improving health through research. 2009. Available at: http://iom.edu/Reports/2009/Beyond-the-HIPAA-Privacy-Rule-Enhancing-Privacy-Improving-Health-Through-Research.aspx, accessed 9/29/12. Citado aquí

15. National Bioethics Advisory Commission. Ethical and policy issues in research involving human participants. Rockville, MD: National Bioethics Advisory Commission, 2001. Citado aquí

16. Bombardier C, Laine L, Reicin A, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. N Engl J Med 2000;343:1520-1528. Citado aquí

17. Curfman GD, Morrissey S, Drazen JM. Expression of concern. N Engl J Med 2005;353:2813-2814. Citado aquí

18. Bresalier RS, Sandler RS, Quan H, et al. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. N Engl J Med 2005;352:1092-1102. Citado aquí

19. Godlee F, Smith J, Marcovitch H. Wakefield's article linking MMR vaccine and autism was fraudulent. BMJ 2011; 342:c7452. Citado aquí

20. Kennedy D. Responding to fraud. Science 2006;314:1353. Citado aquí

21. Office of Research Integrity. Case summaries. Available at: http://ori.hhs.gov/case_summary, accessed 9/29/12. Citado aquí

22. Mello MM, Brennan TA. Due process in investigations of research misconduct. N Engl J Med 2003;349:1280-1286. Citado aquí

23. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. Available at: www.icmje.org/faq_urm.html, accessed 9/29/12. Citado aquí

24. Wislar JS, Flanagin A, Fontanarosa PB, Deangelis CD. Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. BMJ 2011;343:d6128. Citado aquí

25. Browner WS. Authorship. In: Publishing and presenting clinical research, 2nd ed. Philadelphia: Lippincott Williams & Willkins, 2006, 137-144. Citado aquí

26. Lo B, Field M. Conflict of interest in medical research, education, and practice. 2009. Available at: www.iom.edu/Reports/2009/Conflict-of-Interest-in-Medical-Research-Education-and-Practice.aspx, accessed 11/16/11. Citado aquí

27. DeAngelis CD, Fontanarosa PB. Ensuring integrity in industry-sponsored research: primum non nocere, revisited. JAMA 2010;303:1196-1198. Citado aquí

28. DeAngelis CD, Fontanarosa PB. Impugning the integrity of medical science: the adverse effects of industry influence. JAMA 2008;299:1833-1835. Citado aquí

29. Joffe S, Miller FG. Equipoise: asking the right questions for clinical trial design. Nat Rev Clin Oncol 2012;9:230-235. Citado aquí

30. Ellenberg SS, Fleming TR, DeMets DL. Data monitoring committees in clinical trials. Chichester, England: Wiley, 2003. Citado aquí

31. Grady C. Payment of clinical research subjects. J Clin Invest 2005;115:1681-1687. Citado aquí

CAPÍTULO 15 Diseño de cuestionarios, entrevistas y encuestas por Internet

Steven R. Cummings
Michael A. Kohn
Stephen B. Hulley

Mucha de la información utilizada en la investigación clínica se reúne mediante cuestionarios, administrados en papel o con medios electrónicos, o mediante entrevistas. En muchos estudios, la validez de los resultados depende de la calidad de estos instrumentos. En este capítulo se describen los componentes de los cuestionarios y las entrevistas, y se esbozan procedimientos para desarrollarlos.

Los investigadores clínicos disponen de un número rápidamente creciente de opciones para desarrollar encuestas en Internet, como REDCap, una plataforma de gestión de datos basada en Internet y desarrollada por un consorcio de la Universidad Vanderbilt, y productos comerciales, como SurveyMonkey, Zoomerang, Qualtrics y QuesGen. Estos productos ofrecen herramientas en línea para elaborar fácilmente encuestas, y utilidades para el envío automático por correo electrónico a los participantes en el estudio o para remitir la información a la página web del estudio. La transición que se está produciendo entre las encuestas en papel y las encuestas en Internet no ha modificado los principios del diseño de buenos instrumentos: elaborar instrucciones claras y preguntas bien planteadas que permitan obtener respuestas informativas (1).

▪ DISEÑO DE BUENOS INSTRUMENTOS

Preguntas abiertas y cerradas

Existen dos tipos básicos de preguntas, abiertas y cerradas, que sirven para fines algo diferentes. Las preguntas abiertas tienen una utilidad especial cuando es importante oír lo que los participantes tienen que decir con sus propias palabras. Por ejemplo:

¿Qué hábitos cree usted que aumentan la posibilidad de que una persona sufra una trombosis cerebral?

Las preguntas abiertas dejan que la persona responda libremente con menos límites impuestos por el investigador. Permiten que los participantes comuniquen más información de la que es posible con una lista de respuestas, aunque las respuestas puedan ser menos completas. Un importante inconveniente es que las preguntas abiertas suelen precisar métodos cualitativos o sistemas especiales (como diccionarios de códigos para síntomas y acontecimientos adversos) para codificar y analizar las respuestas; esto conlleva más tiempo que introducir respuestas a preguntas cerradas, y puede necesitar juicios subjetivos. Las preguntas abiertas suelen usarse en las fases de exploración del diseño de las preguntas, porque ayudan al investigador a comprender un concepto tal y como lo expresan los participantes que responden. Las frases y palabras que usan los participantes pueden constituir la base para preguntas cerradas que piden a los participantes que elijan entre dos o más respuestas seleccionadas previamente:

¿Cuál de las conductas siguientes cree que aumenta la posibilidad de sufrir una trombosis cerebral?

(Marque todas las que correspondan.)

□ Fumar

□ Tener sobrepeso

□ Estrés

□ Beber alcohol

Como las preguntas cerradas proporcionan una lista de alternativas posibles entre las que el participante puede elegir, son más rápidas y más fáciles de responder, y las respuestas son más fáciles de tabular y analizar. Además, la lista de posibles respuestas a menudo ayuda a aclarar el significado de la pregunta, y estas preguntas son adecuadas para su uso en escalas con múltiples preguntas para obtener una sola puntuación.

Las preguntas cerradas también tienen varios inconvenientes. Conducen a los participantes en determinadas direcciones y no les permiten expresar sus propias, y posiblemente más exactas, respuestas. El grupo de respuestas puede no ser exhaustivo (no incluir todas las opciones posibles, p. ej., «actividad sexual» o «la sal de la dieta»). Una solución es incluir una opción como «Otros (por favor, especifíquelo)» o «Ninguna de las anteriores». Cuando se desea una sola respuesta, debe indicarse así al participante, y el grupo de posibles respuestas debe ser mutuamente excluyente (no deben superponerse las categorías), para asegurar la claridad y la parquedad [1] .

Cuando la pregunta permite más de una respuesta, lo mejor no es decir al participante «Todas las que procedan». Esto no obliga al participante a considerar todas las posibles respuestas, y una pregunta que se pasa por alto puede representar una respuesta que no

corresponde o una pregunta que se ha pasado por alto. Es mejor pedir que se marque cada posible respuesta con un «sí» o un «no», como en el ejemplo.

¿Cuál de los siguientes cree que aumenta la posibilidad de sufrir una trombosis cerebral?

	Sí	No	No sabe
Fumar	○	○	○
Tener sobrepeso	○	○	○
Estrés	○	○	○
Beber alcohol	○	○	○

La escala analógica visual (EAV) es otra opción para registrar respuestas a preguntas cerradas usando líneas u otros gráficos. Se pide al participante que marque una línea en un punto, a lo largo del recorrido continuo de un extremo al otro, que represente mejor su respuesta. Es importante que las palabras de cada extremo describan los valores más extremos para cada tema de interés. He aquí un ejemplo sobre la intensidad del dolor:

Por favor, marque el punto de esta línea que describa mejor la intensidad de su dolor, en general, a lo largo de la última semana.

Ninguno Insoportable

Para facilitar la medición, las líneas suelen tener una longitud de 10 cm, y la puntuación es la distancia, en centímetros, desde el extremo más bajo. Se puede ver un ejemplo de una EAV en Internet en la página web: http://www.epibiostat.ucsf.edu/dcr/.

Las EAV son atractivas porque valoran características en una escala continua; pueden ser más sensibles a cambios pequeños que las puntuaciones que se basan en listas categóricas de adjetivos. Muchas de las herramientas para elaborar encuestas en Internet, como REDCap, Qualtrics y Ques-Gen, permiten el uso de EAV.

Formato

En los cuestionarios es habitual describir brevemente al comienzo el objetivo del estudio y el modo en que se usarán los datos. Una información similar se presenta al principio de una entrevista, como parte de la obtención del consentimiento. Para asegurar que se obtendrán respuestas exactas y normalizadas, todos los instrumentos deben tener instrucciones que especifiquen cómo deben rellenarse. Esto es así no solo en los cuestionarios autoadministrados, sino también en los formularios que usan los entrevistadores para apuntar las respuestas.

A veces es útil proporcionar un ejemplo de cómo rellenar una pregunta, utilizando una pregunta sencilla que se responda fácilmente:

Instrucciones para rellenar un cuestionario que valora el aporte dietético.

Estas preguntas se refieren a sus hábitos alimentarios habituales durante los últimos 12 meses. Marque, por favor, la cantidad habitual que se sirve y escriba con qué frecuencia consume cada alimento en los recuadros junto a cada uno de ellos.

Por ejemplo, si toma un vaso de tamaño medio de zumo de manzana unas tres veces a la semana, debería responder:

Zumo de manzana ○ Pequeño (125 ml) [3] veces por ○ Día

◎ Medio (250 ml)	◎ Semana
○ Grande (375 ml)	○ Mes
	○ Año

Para mejorar el flujo del instrumento, deben agruparse juntas todas las preguntas referentes a áreas importantes, y presentarse con encabezamientos con frases descriptivas cortas. Para estimular al participante a empezar el proceso de responder a las preguntas, es útil empezar con preguntas emocionalmente neutras, como el nombre y la información de contacto. Las preguntas muy delicadas sobre los ingresos económicos o la función sexual generalmente se sitúan al final del instrumento. Para cada pregunta o grupo de ellas con un formato que difiere del de otras preguntas del instrumento, las instrucciones deben indicar claramente cómo responder.

Si las instrucciones incluyen diferentes estructuras cronológicas, a veces resulta útil repetir la estructura en la parte superior de cada nuevo grupo de preguntas. Por ejemplo, preguntas como:

¿Con qué frecuencia ha acudido al médico en el último año?

Durante el último año, ¿cuántas veces ha estado como paciente en un servicio de urgencias?

¿Cuántas veces ingresó en el hospital durante el último año?

pueden acortarse y ordenarse así:

Durante el último año, ¿cuántas veces…

- ha acudido al médico?
- ha estado como paciente en un servicio de urgencias?
- ha ingresado en un hospital?

Para formularios en papel, el diseño visual de los instrumentos debe ser lo más fácil posible para que quienes los cumplimentan, ya sean participantes en el estudio o el personal de la investigación, contesten a todas las preguntas en el orden correcto. Si la disposición es demasiado compleja, los participantes o los entrevistadores pueden saltarse preguntas, dar una información errónea e incluso negarse a cumplimentarlo. Una disposición clara y con abundante espacio es más atractiva y fácil de usar que una que esté apiñada. Aunque los investigadores suponen a veces que un cuestionario parecerá más corto si tiene menos páginas, la tarea puede ser más difícil cuando se incluyen más preguntas en una página. Las escalas de respuesta deben tener espacios suficientemente amplios para que sea fácil marcar o encerrar en un círculo el número correcto, sin incluir accidentalmente la marca en la respuesta «anterior» o «posterior». Cuando se incluye una pregunta abierta, el espacio para responder debe tener el tamaño suficiente para permitir que los que tienen una escritura grande puedan escribir cómodamente en el espacio. Las personas con problemas visuales, entre ellos muchos ancianos, agradecerán un cuerpo de letra grande y un contraste elevado (negro sobre blanco).

Las posibles respuestas a las preguntas cerradas deben alinearse verticalmente e ir precedidas por recuadros o paréntesis para marcar, o por números para encerrar en un círculo, en lugar de por espacios en blanco:

¿Cuántos medicamentos diferentes toma cada día? (Marque una.)

○ Ninguno

○ 1-2

○ 3-4

○ 5-6

○ 7 o más

Observe que estas opciones de respuesta son exhaustivas y mutuamente excluyentes.

A veces el investigador puede desear hacer un seguimiento de determinadas respuestas con preguntas más detalladas. Lo mejor para ello es la pregunta ramificada. Las respuestas a la pregunta inicial, a menudo denominada «detectora», determinan si están dirigidas a responder otras preguntas o a saltar a preguntas posteriores. Por ejemplo:

¿Le han dicho alguna vez que tiene la tensión alta?

○ Sí →

¿Qué edad tenía la primera vez que le dijeron que tenía la tensión alta?

○ No

↓ _ _ años de edad

Vaya a la pregunta 11

Las preguntas ramificadas ahorran tiempo y permiten que los participantes eviten preguntas irrelevantes o redundantes. Para dirigir al participante hacia la siguiente respuesta adecuada se usan flechas que señalan la respuesta a las preguntas de seguimiento, y se incluyen instrucciones como «Vaya a la pregunta 11» (apéndice 15).

Las encuestas en Internet generalmente son más claras y fáciles para los participantes, porque incorporan lógica de exclusión. Un hombre no verá una pregunta sobre embarazos y solo llegará a la pregunta sobre el número de cajetillas/año si ha respondido «sí» a la pregunta sobre tabaquismo. (V. www.epibiostat.ucsf.edu/dcr/.) Sin embargo, la lógica de exclusión se debe validar cuidadosamente durante la fase de verificación previa del estudio. Una lógica de exclusión compleja puede llevar a callejones sin salida y a preguntas «huérfanas» a las que nunca se llega. Un buen diseño, que tenga en cuenta también a los participantes con problemas visuales [2] , tiene la misma importancia para los formularios en Internet que para los formularios en papel.

Redacción

Cada palabra de un cuestionario puede influir en la validez y la reproducibilidad de las respuestas. El objetivo es elaborar preguntas que sean sencillas y no ambiguas, y que permitan obtener respuestas exactas y sinceras sin que el participante se sienta turbado u ofendido.

- Claridad. Haga preguntas lo más claras y específicas posibles. Se prefieren términos concretos mejor que abstractos. Por ejemplo, preguntar «¿Cuánto ejercicio suele realizar?» es menos claro que «Durante una semana típica, ¿cuántas horas pasa andando a buen ritmo?»

- Sencillez. Utilice palabras y una gramática sencillas y de uso habitual, y evite los términos técnicos y la jerga. Por ejemplo, es más claro preguntar por «medicinas que puede comprar sin una receta del médico» que hacerlo por «medicamentos de libre dispensación».

- Neutralidad. Hay que evitar términos «cargados» y estereotipos que sugieran una respuesta deseable. Preguntar «Durante el último mes, ¿con qué frecuencia ha consumido alcohol en exceso?» puede hacer que los participantes no admitan que consumen alcohol en exceso. «Durante el último mes, ¿con qué frecuencia ha consumido más de cinco bebidas alcohólicas al día?» es una pregunta más real, menos subjetiva y menos ambigua.

A veces es útil establecer un tono que permita al participante admitir conductas y actitudes que puedan considerarse no deseables. Por ejemplo, cuando se pregunta sobre el cumplimiento de los pacientes con los fármacos recetados, un entrevistador o un cuestionario pueden utilizar una introducción: «A veces las personas olvidan tomar los medicamentos que el médico les ha recetado. ¿Le ha sucedido alguna vez a usted?» Sin embargo, este enunciado puede plantear dificultades; es importante autorizar a los participantes a admitir ciertos comportamientos sin estimularles a que lo exageren.

Obtener información sobre temas posiblemente delicados, como la conducta sexual o los ingresos económicos, puede ser especialmente difícil. Algunas personas se sienten más cómodas respondiendo a este tipo de preguntas mediante cuestionarios autoadministrados que en entrevistas, pero un entrevistador experimentado puede, a veces, lograr respuestas abiertas y sinceras. Puede ser útil poner las preguntas

posiblemente embarazosas en una tarjeta, de modo que solo sea necesario señalar una respuesta para responder.

Establecer la estructura cronológica

Para medir la frecuencia de la conducta, es esencial describirla al participante mediante alguna unidad de tiempo. Si la conducta suele ser la misma día tras día, como tomar un comprimido de un diurético por la mañana, la pregunta puede ser muy sencilla: «¿Cuántos comprimidos toma al día?»

Muchas conductas varían de un día a otro, según las estaciones o de año en año. Para medirlas, el investigador debe decidir primero cuál es el aspecto más importante de esa conducta en relación con el estudio: el promedio o los extremos. Por ejemplo, un estudio del efecto del alcohol sobre el riesgo de sufrir una enfermedad cardiovascular puede precisar una determinación del consumo medio a lo largo del tiempo, pero un estudio de la influencia del alcohol en la incidencia de caídas puede necesitar saber con qué frecuencia el participante bebe alcohol suficiente como para emborracharse.

Las preguntas sobre una conducta media pueden hacerse de dos formas: preguntando sobre la conducta «habitual» o «típica», o contando conductas reales durante un período de tiempo. Por ejemplo, un investigador puede determinar el consumo medio de cerveza pidiendo a los participantes que calculen su consumo habitual:

¿Alrededor de cuántas cervezas toma durante 1 semana normal (una cerveza equivale a una lata o una botella de 340 ml, o un vaso grande)?

[___] cervezas a la semana

Este formato es sencillo y corto. Supone, sin embargo, que los participantes pueden promediar con exactitud su conducta mediante un cálculo sencillo. Como los patrones de consumo con frecuencia varían notablemente incluso en intervalos cortos, puede que el participante tenga dificultad en decidir qué es una semana normal. Ante preguntas que interrogan sobre conductas habituales o típicas, las personas a menudo cuentan las cosas que hacen más habitualmente e ignoran los extremos. Si se pregunta sobre el consumo en días típicos, por ejemplo, se infravalorará el consumo de alcohol si quien responde suele beber grandes cantidades en el fin de semana.

Un método alternativo sería cuantificar la exposición durante un tiempo determinado:

Durante los últimos 7 días, ¿cuántas cervezas se ha tomado (una cerveza equivale a una lata o una botella de 340 ml, o un vaso grande)?

[___] cervezas en los últimos 7 días

El objetivo es preguntar por el período de tiempo último más corto que represente de modo exacto las características de todo el período de interés para la pregunta de la investigación. La mejor duración de ese período depende de la característica. Por ejemplo, los patrones de sueño pueden variar considerablemente de un día a otro, pero las preguntas sobre hábitos de sueño durante la última semana pueden representar adecuadamente los patrones de sueño durante el año entero. Por otro lado, la frecuencia de las relaciones sexuales sin protección puede variar mucho de una semana a otra, por lo que las preguntas sobre esta cuestión deben abarcar intervalos más prolongados.

El uso de diarios puede ser un método más exacto para seguir la trayectoria de acontecimientos, conductas o síntomas que suceden de forma episódica (como las caídas) o que varían de un día a otro (como la hemorragia vaginal). Esto puede ser útil cuando la secuencia cronológica o la duración de un episodio son importantes, o cuando se olvida fácilmente su aparición. Los participantes pueden introducir esos datos en aparatos electrónicos, y el método permite que el investigador calcule una puntuación diaria media del fenómeno o la conducta que se está evaluando. Sin embargo, este método requiere tiempo de los participantes y puede hacer que se pierdan más datos que con las preguntas retrospectivas. El uso de diarios supone que el período evaluado era típico, y que el conocimiento de uno mismo necesario para usar los diarios no ha alterado la conducta que se registra de manera importante.

Evitar las dificultades

- Preguntas de doble efecto. Cada pregunta solo debe contener un concepto. Considere esta pregunta destinada a valorar el consumo de cafeína: «¿Cuántas tazas de café o té toma a lo largo del día?» El café contiene mucha más cafeína que el té, y es diferente en otros aspectos, por lo que una respuesta que combine ambas bebidas no es tan precisa como debería ser. Cuando una pregunta intenta valorar dos cosas al mismo tiempo, es mejor dividirla en dos preguntas separadas. «(1) ¿Cuántas tazas de café toma a lo largo de un día normal?» y «(2) ¿Cuántas tazas de té toma a lo largo de un día normal?»

- Suposiciones ocultas. A veces las preguntas conllevan suposiciones que no pueden aplicarse a todas las personas que participan en el estudio. Por ejemplo, una pregunta habitual sobre la depresión se refiere a la frecuencia en la última semana: «Sentí que no

podía superar la tristeza incluso con la ayuda de mi familia.» Esto supone que quienes responden tienen familia y le piden apoyo emocional; para los que no tengan familia o no le pidan ayuda, es difícil contestar a la pregunta.

- La pregunta y las opciones de respuesta no encajan. Es importante que la pregunta concuerde con las opciones de respuesta, una tarea que parece sencilla, pero que a menudo se realiza de forma incorrecta. Por ejemplo, la pregunta «¿Ha sentido dolor en la última semana?» no se debe relacionar con las opciones de respuesta «nunca», «rara vez», «a menudo», «muy a menudo». (La pregunta debe cambiarse a «¿Con qué frecuencia ha sentido dolor en la última semana?», o la respuesta debe cambiare a «sí» o «no».) Otro problema frecuente surge cuando, ante preguntas sobre intensidad, se proporcionan opciones de acuerdo/desacuerdo. Por ejemplo, a un participante se le puede presentar, para que la responda, la frase «A veces estoy deprimido», y luego pedirle que responda «estoy de acuerdo» o «no estoy de acuerdo». No estar de acuerdo con esa frase podría significar que la persona está a menudo deprimida o que nunca lo está. Suele ser mucho más claro usar una pregunta sencilla sobre la frecuencia con la que la persona se siente deprimida y relacionarla con opciones sobre frecuencia (nunca, a veces, a menudo).

Escalas y puntuaciones para medir variables abstractas

Es difícil valorar cuantitativamente conceptos abstractos, como la calidad de vida, con preguntas sencillas. Por lo tanto, las características abstractas suelen medirse generando puntuaciones a partir de una serie de preguntas que se organizan en una escala (2 , 3).

Usar múltiples preguntas para valorar un concepto puede tener otras ventajas sobre las preguntas únicas o sobre varias preguntas que se plantean de formas diferentes y que no pueden combinarse. En comparación con otros métodos, las escalas con múltiples preguntas pueden aumentar el espectro de posibles respuestas (p. ej., una escala de múltiples preguntas sobre calidad de vida podría generar puntuaciones que oscilen entre 1 y 100, mientras que una sola pregunta podría dar cuatro o cinco respuestas desde «mala» hasta «excelente»). Un inconveniente de las escalas de múltiples preguntas es que producen resultados (calidad de vida = 46,2) que pueden ser difíciles de entender intuitivamente.

Las escalas de Likert suelen usarse para cuantificar actitudes, comportamientos y dominios de calidad de vida relacionada con la salud. Estas escalas proporcionan a los participantes una lista de afirmaciones o preguntas, y les piden que elijan la respuesta que mejor represente el grado de su respuesta. A cada respuesta se asigna un número de puntos. Por ejemplo, considere un cuestionario para medir la intensidad de la opinión de una persona de que una dieta rica en fruta y verduras mejora la salud:

Para cada pregunta, rodee el número que representa mejor su opinión:

		Muy de acuerdo	De acuerdo	Neutral	En desacuerdo	Muy en desacuerdo
a.	Comer más fruta y verduras reduce el riesgo de enfermedades del corazón	1	2	3	4	5
b.	Los vegetarianos están más sanos que las personas que incluyen carne en la dieta	1	2	3	4	5
c.	El aumento de la ingesta de frutas y verduras retrasa el envejecimiento	1	2	3	4	5

Un investigador puede calcular una puntuación total para las respuestas de una persona simplemente sumando la puntuación de cada pregunta, o promediando los puntos de todas las preguntas que se hayan contestado. Por ejemplo, una persona que respondió que consumir más frutas y verduras reduce el riesgo de enfermedades del corazón (un punto) y que los vegetarianos están más sanos que las personas que incluyen carne en la dieta (un punto), pero que estuvo en desacuerdo en que el aumento de la ingesta de frutas y verduras retrasa el envejecimiento (cuatro puntos), tendría una puntuación total de seis. Simplemente sumar o promediar las puntuaciones de las preguntas supone que todas las preguntas tienen el mismo peso y que cada pregunta está midiendo la misma característica general.

La consistencia interna de una escala puede verificarse estadísticamente usando medidas como el estadístico *a*de Cronbach (4), que valora la consistencia global de una escala. El estadístico *a* de Cronbach se calcula a partir de las correlaciones entre las puntuaciones de preguntas individuales. Se considera que los valores de esta medida que sean superiores a 0,80 son excelentes, y por debajo de 0,50 son

inaceptables. Valores bajos de consistencia interna indican que algunas de las preguntas individuales pueden estar midiendo características diferentes.

Creación de nuevas escalas

Cuando un investigador tiene que medir una característica para la que no hay ningún cuestionario ni abordaje de entrevista estandarizado, puede ser necesario elaborar un nuevo instrumento o escala. La labor puede oscilar desde la creación de una sola pregunta nueva sobre una variable poco importante en un estudio pequeño, hasta el desarrollo y el estudio de una nueva escala con múltiples preguntas para medir la variable de respuesta principal en un estudio multicéntrico. En el extremo más sencillo de este espectro, el investigador puede usar su buen juicio y los principios básicos de la escritura para desarrollar una pregunta que se pueda verificar antes de aplicarla en el estudio para asegurarse de que está clara y de que produce respuestas adecuadas. En el otro extremo, el desarrollo de un nuevo instrumento para medir un concepto importante puede necesitar un método sistemático que puede durar años desde el borrador inicial hasta el producto final.

El último proceso, a menudo, empieza generando posibles preguntas para el instrumento a partir de entrevistas con personas y grupos de discusión (pequeños grupos de personas importantes para la pregunta de investigación, a los que se invita a pasar 1 o 2 h comentando temas específicos sobre el estudio con un director del grupo). Después se elabora un instrumento, a lo que sigue la revisión crítica por revisores externos, mentores y expertos. Los investigadores realizan entonces la secuencia iterativa de evaluación preliminar, revisión, reducción y validación que se describe en la sección siguiente (y que se ilustra en el ejemplo 15-1).

EJEMPLO 15-1. Desarrollo de un nuevo instrumento con múltiples preguntas

El *National Eye Institute Visual Function Questionnaire* ejemplifica el desarrollo y la evaluación cuidadosos de un instrumento con múltiples preguntas. Mangione y cols. dedicaron varios años a crear y evaluar la escala, porque se pretendía que sirviera como medición principal de la variable de respuesta de muchos estudios de afecciones oculares (5 , 6). Empezaron entrevistando a pacientes con enfermedades oculares sobre los modos en que estas afecciones afectaban a sus vidas. A continuación, organizaron grupos de discusión de pacientes con las enfermedades, y analizaron transcripciones de esas sesiones para elegir preguntas pertinentes y opciones de respuesta. Elaboraron y realizaron la evaluación preliminar de un extenso cuestionario que se administró a cientos de participantes en varios estudios. Usaron datos de estos estudios para identificar las preguntas que conformaban la mayor contribución a la variación de las puntuaciones de unas personas a otras, y para acortar el cuestionario de 51 a 25 preguntas.

Debido a que la creación y la validación de instrumentos de múltiples preguntas es laboriosa, solo deben realizarse, generalmente, para variables que son fundamentales para un estudio, y cuando las medidas existentes no son adecuadas ni apropiadas para las personas a las que se va a incluir en el estudio.

Volver al principio

▪ PASOS PARA REUNIR LOS INSTRUMENTOS PARA EL ESTUDIO

Elaboración de una lista de variables

Antes de diseñar una entrevista o cuestionario, elabore una lista detallada de la información que quiere recoger y los conceptos que se deben medir en el estudio. Plantéese la elaboración de una lista con la contribución de cada uno de los elementos (p. ej., variables predictivas, variables de respuesta y posibles factores de confusión) a la respuesta a las preguntas de la investigación principal.

Prefiera medidas ya existentes cuando sea posible

Reúna un archivo de preguntas o instrumentos disponibles para medir cada variable. Cuando haya varios métodos alternativos, cree un archivo electrónico para cada variable que vaya a medir, y luego encuentre y archive copias de posibles preguntas o instrumentos para cada elemento. Es importante usar los mejores instrumentos posibles para medir las principales variables predictivas y de respuesta de un estudio, por lo que la mayor parte del esfuerzo de recopilación de instrumentos alternativos debe centrarse en estas variables principales.

Comience recopilando instrumentos de otros investigadores que hayan realizado estudios que hayan incluido las mediciones de interés. Los cuestionarios existentes y la información sobre su validez, consistencia interna y fiabilidad se pueden encontrar en la sección de metodología de los trabajos publicados, y haciendo una búsqueda en Internet con palabras clave como «cuestionarios de resultados en salud» (*health outcomes questionnaires*).

Tomar prestados instrumentos de otros estudios tiene la ventaja de que se ahorra tiempo y permite que se comparen los resultados de distintos estudios. Lo ideal es usar los instrumentos sin modificaciones. No obstante, si algunas de las preguntas no son adecuadas (como

puede suceder cuando un cuestionario desarrollado para un grupo cultural se aplica en un entorno diferente), puede ser necesario eliminar, cambiar o añadir algunas preguntas.

Si un instrumento establecido es demasiado largo, el investigador puede contactar con los autores que lo desarrollaron para ver si cuentan con versiones más abreviadas. Si se eliminan preguntas de escalas ya establecidas, se corre el riesgo de cambiar el significado de las puntuaciones, y se ponen en peligro las comparaciones de los hallazgos con los resultados de los estudios en los que se usó la escala intacta. Si se abrevia una escala, pueden reducirse también su reproducibilidad o su sensibilidad para detectar cambios. Sin embargo, a veces, puede aceptarse la eliminación de secciones o «subescalas» que no son esenciales para el estudio si se dejan las demás partes intactas.

Prepare un nuevo instrumento, si es necesario

El primer borrador del instrumento debe tener un alcance amplio, incluyendo más preguntas sobre el tema de las que finalmente se incluirán en el instrumento. El investigador debe leer con atención el primer borrador, intentando responder a todas las preguntas como si fuera un participante e intentando imaginar las formas de interpretar erróneamente las preguntas. El objetivo es identificar palabras o frases que podrían ser confusas o llevar a malentendidos, encontrar palabras abstractas o jerga que se pudieran traducir en términos más sencillos y concretos, y detectar preguntas complejas que se pueden desglosar en dos o más preguntas. Se debe pedir a compañeros y expertos en el diseño de cuestionarios que revisen el instrumento, teniendo en consideración tanto el contenido de las preguntas como la claridad.

Revisión y abreviación de la serie de instrumentos para el estudio

Los estudios suelen recopilar más datos de los que se analizarán. Las entrevistas, cuestionarios y exámenes prolongados pueden cansar a los participantes y, por lo tanto, pueden reducir la exactitud y la reproducibilidad de sus respuestas. Habitualmente, es mejor resistir a la tentación de incluir preguntas y medidas adicionales «por si» podrían producir datos interesantes. Las preguntas que no son esenciales para responder a la pregunta de la investigación principal aumentan la magnitud del esfuerzo implicado en la obtención, introducción, limpieza y análisis de los datos. El tiempo que se dedica a los datos innecesarios o de un valor escaso puede distraer de otros esfuerzos y disminuir la calidad y la productividad generales del estudio.

Para decidir si un concepto es esencial, el investigador puede planear por adelantado el análisis y la descripción de los resultados del estudio. Esbozar las tablas finales ayudará a asegurar que están incluidas todas las variables necesarias y a identificar las que son menos importantes. Una vez que se ha realizado, aquí hay una máxima para decidir qué elementos se deben incluir: en caso de duda, déjelo fuera.

Estudio preliminar

Realice un estudio preliminar de la claridad del instrumento y del tiempo que se tarda en rellenar. Para las mediciones esenciales, los grandes estudios piloto pueden ser útiles para observar si cada pregunta produce un espectro adecuado de respuestas, y para estudiar la validez y la reproducibilidad del instrumento.

Validación

Puede valorarse la validez (un aspecto de la exactitud) y la reproducibilidad (precisión) de los cuestionarios y las entrevistas del mismo modo que cualquier otro tipo de medición (cap. 4). El proceso se inicia escogiendo preguntas que tienen validez aparente, el juicio subjetivo, pero importante, de que las preguntas valoran las características de interés, y continúa con el intento de determinar la validez de contenido y la validez de constructo. Cuando sea posible, pueden compararse los nuevos instrumentos con métodos de referencia establecidos para medir la situación de interés. Finalmente, puede valorarse la validez predictiva de un instrumento relacionando las mediciones con futuras respuestas.

Si se pretende que un instrumento mida cambios, puede probarse su sensibilidad aplicándolo a los pacientes antes y después de recibir tratamientos que se consideran eficaces por otras mediciones. Por ejemplo, un nuevo instrumento diseñado para medir la calidad de vida en personas con alteración de la agudeza visual podría incluir preguntas que tengan validez aparente («¿Puede leer un periódico sin gafas ni lentes de contacto?»). Pueden compararse las respuestas con las respuestas a un instrumento validado existente (ejemplo 15-1) en pacientes con cataratas avanzadas y en personas con exploraciones oculares normales. Puede analizarse la sensibilidad al cambio del instrumento comparando respuestas de los pacientes con cataratas antes y después de la cirugía. Sin embargo, el proceso de validación de nuevos instrumentos es laborioso y costoso, y está justificado solo si los instrumentos existentes no son adecuados para la pregunta de la investigación o la población que se va a estudiar.

Volver al principio

▪ ADMINISTRACIÓN DE LOS INSTRUMENTOS

Cuestionarios y entrevistas

Hay dos métodos básicos para recopilar datos sobre actitudes, conductas, conocimientos, salud y antecedentes personales. Los cuestionarios son instrumentos que los participantes rellenan por sí solos, y las entrevistas son las que administra verbalmente un entrevistador. Cada uno de los métodos tiene sus ventajas y sus inconvenientes.

Los cuestionarios suelen ser un modo más eficaz y uniforme de plantear preguntas sencillas, como la edad o los hábitos de consumo de tabaco. Son más baratos que las entrevistas, porque precisan menos tiempo del personal de investigación y se normalizan con más facilidad. Las entrevistas suelen ser un método mejor para recopilar respuestas a preguntas más complejas que requieren una explicación u orientación, y los entrevistadores pueden asegurarse de que las respuestas sean completas. Pueden ser necesarias las entrevistas cuando los participantes tienen alteraciones de la capacidad de leer y entender las preguntas. Sin embargo, son más costosas y laboriosas, y las respuestas pueden verse influidas por la relación entre el entrevistador y el entrevistado.

Ambos tipos de instrumentos pueden normalizarse, pero las entrevistas inevitablemente se administran con alguna diferencia de unas veces a otras. Estos dos métodos de obtención de información pueden producir errores causados por una memoria imperfecta; ambos se ven afectados también, aunque no necesariamente en el mismo grado, por la tendencia de los participantes a dar respuestas socialmente aceptables.

Entrevista

La habilidad del entrevistador puede tener un gran efecto sobre la calidad de las respuestas. Para aumentar al máximo la reproducibilidad, debe normalizarse el procedimiento de una entrevista a otra, con un enunciado uniforme de las preguntas y con señales no verbales uniformes durante la entrevista. Los entrevistadores deben intentar no introducir sus propios sesgos en las respuestas cambiando las palabras o el tono de su voz. Para que el entrevistador lea cómodamente las preguntas palabra por palabra, la entrevista debe redactarse en un lenguaje similar al habla común. Las preguntas que parecen ampulosas o no naturales cuando se dicen en voz alta estimularán a los entrevistadores a improvisar su propio modo, más natural, pero menos normalizado, de hacer las preguntas.

A veces es necesario hacer un seguimiento de las respuestas del participante para estimularle a que dé una respuesta adecuada o para aclarar el significado de otra. Este «sondeo» también puede normalizarse redactando frases estandarizadas en los márgenes o debajo del texto de cada pregunta. Para una pregunta sobre cuántas tazas de café consumen los participantes en un día normal, algunos podrían responder «No estoy seguro; varía de un día a otro». El instrumento podría incluir la siguiente afirmación de sondeo: «Hágalo lo mejor que pueda; dígame aproximadamente cuántas toma en un día normal».

Las entrevistas pueden realizarse en persona o por teléfono. La entrevista telefónica asistida por ordenador (ETAO) es una técnica de encuesta telefónica en la que el entrevistador sigue un guion y el ordenador facilita la recogida y la edición de los datos. Los sistemas de respuesta interactiva a la voz (RIV) sustituyen al entrevistador con preguntas generadas por ordenador que recogen las respuestas de los participantes mediante un teclado auxiliar o reconocimiento de la voz (7). Sin embargo, pueden ser necesarias las entrevistas en persona si el estudio precisa la observación directa de los participantes o la realización de exploraciones físicas, o si los posibles participantes no tienen teléfono (p. ej., indigentes).

Métodos para administrar los cuestionarios

Se pueden entregar los cuestionarios a los participantes en persona o pueden enviarse por correo postal, correo electrónico o a través de una página web. Si se distribuyen en persona, el investigador puede explicar las instrucciones antes de que el participante empiece a responder a las preguntas. Cuando la investigación necesita que el participante acuda al centro de investigación para la realización de pruebas, pueden enviársele los cuestionarios antes de que acuda a la cita, y allí puede comprobarse que las respuestas están completas antes de que el participante se vaya.

Los cuestionarios por correo electrónico tienen algunas ventajas sobre los que se envían por correo postal. Aunque solo pueden mandarse a los participantes que tengan acceso y estén familiarizados con Internet, los cuestionarios enviados por correo electrónico son una forma sencilla de obtener datos que se puedan introducir directamente en bases de datos.

Los cuestionarios de páginas web o con dispositivos electrónicos portátiles se están utilizando cada vez más como abordajes eficientes y económicos para obtener información en encuestas sanitarias (8). Estos métodos permiten obtener datos muy claros, porque las respuestas pueden comprobarse automáticamente por si presentan valores ausentes o extremos, pueden indicarse los errores al participante, y las respuestas se aceptan solo tras corregirse los errores.

Volver al principio

▪ CONSIDERACIÓN DE MEDICIONES DIRECTAS

Los avances en los instrumentos de medición y los análisis biológicos están creando alternativas a los cuestionarios y las entrevistas para medir muchas enfermedades y exposiciones frecuentes. Por ejemplo, la medición directa de la actividad física llevando pequeños acelerómetros ofrece una información más objetiva y precisa de la actividad total, los patrones de la actigrafía y el gasto energético que los cuestionarios sobre la actividad física (9). Los sensores que se llevan por la noche pueden medir con exactitud la cantidad y la calidad del sueño (10). La medición de las concentraciones sanguíneas de nutrientes como la vitamina D ofrece una medición más exacta de la exposición al nutriente que preguntar por el consumo de alimentos que contienen vitamina D. Los investigadores deben estar alerta a las nuevas tecnologías, muchas veces con dispositivos electrónicos inalámbricos, que miden directamente características que previamente se evaluaban tan solo indirectamente con cuestionarios y entrevistas.

Volver al principio

▪ RESUMEN

- En muchos estudios clínicos, la calidad de los resultados depende de la calidad y de la idoneidad de los cuestionarios y las entrevistas. Los investigadores deben asegurarse de que los instrumentos sean lo más válidos y reproducibles posible antes de que se inicie el estudio.
- Las preguntas abiertas permiten que los participantes respondan sin limitaciones impuestas por el investigador, y las preguntas cerradas son más fáciles de responder y de analizar. Las opciones de respuesta a una pregunta cerrada deben ser exhaustivas y mutuamente excluyentes.
- Las preguntas deben ser claras, sencillas, neutras y adecuadas para la población que se va a estudiar. Los investigadores deben examinar posibles preguntas desde el punto de vista de los posibles participantes, buscando los términos ambiguos y los errores habituales, como las preguntas ambiguas, las suposiciones ocultas y las opciones de respuesta que no concuerdan con la pregunta.
- Los cuestionarios deben ser fáciles de leer, y las preguntas de la entrevista deben ser cómodas de leer en voz alta. El formato debe encajar en el método para la introducción electrónica de los datos, y debe ser espacioso y estar despejado.
- Para medir variables abstractas, como actitudes o calidad de vida, pueden combinarse las preguntas en escalas múltiples para generar una puntuación total. Cuando se usan estas puntuaciones se supone que la pregunta mide una sola característica y que las respuestas tienen consistencia interna.
- Un investigador debe buscar y utilizar instrumentos existentes que se sabe que producen resultados válidos y fiables. Cuando es necesario modificar medidas existentes o idear una nueva, el investigador debe empezar por recopilar las medidas existentes para usarlas como posibles modelos y fuentes de ideas.
- Todos los instrumentos que se van a usar en un estudio deben evaluarse antes de que empiece el estudio. Con los instrumentos nuevos, en pequeñas pruebas preliminares iniciales se puede mejorar la claridad de las preguntas y de las instrucciones; con estudios piloto posteriores, de mayor tamaño, se pueden evaluar y mejorar el intervalo de medición, la reproducibilidad y la validez del nuevo instrumento.
- Los cuestionarios autoadministrados son más económicos que las entrevistas, pueden normalizarse con más facilidad, y su privacidad puede aumentar la validez de las respuestas. Por otro lado, las entrevistas pueden garantizar que se obtengan respuestas más complejas y aumentar la validez mediante una mejor comprensión.
- La administración de instrumentos por entrevistas telefónicas asistidas por ordenador, correo electrónico, dispositivos electrónicos portátiles o a través de la página web del estudio pueden mejorar la eficacia de un estudio.

Volver al principio

[1]Para los formularios en Internet, la convención es presentar las opciones mutuamente excluyentes como botones (círculos), y utilizar casillas de verificación (cuadrados) para las respuestas a preguntas de tipo de «todos los que procedan».

[2]Los proveedores comerciales de herramientas para elaborar encuestas en Internet prestan mucha atención a los problemas de legibilidad, en parte porque la sección 508 de la Ley de Rehabilitación de 1973 exige que las agencias federales velen por que sus formularios electrónicos sean accesibles para personas con discapacidades. La mayoría de los proveedores comerciales tienen la certificación de que cumplen la sección 508.

APÉNDICE 15Ejemplo de un cuestionario sobre tabaquismo
NA

Las siguientes preguntas se han obtenido de un cuestionario en papel autoadministrado usado en nuestro estudio de las fracturas osteoporóticas (*Study of Osteoporotic Fractures*). Obsérvese que las preguntas ramificadas van seguidas por flechas que dirigen al participante a la siguiente pregunta adecuada, y que la disposición está despejada, con todas las respuestas alineadas a la izquierda en todas las áreas sucesivas. Se puede ver un enlace a una versión en Internet de este ejemplo en www.epibiostat.ucsf.edu/dcr/.

1. ¿Ha fumado al menos 100 cigarrillos en toda su vida?

7. ¿Ha vivido alguna vez al menos 1 año en la misma casa con alguien que fumaba cigarrillos habitualmente?

BIBLIOGRAFÍA

1. Iarossi G. The power of survey design: a user guide for managing surveys, interpreting results, and influencing respondents. Washington, DC: World Bank, 2006. Available at: https://openknowledge.worldbank.org/bitstream/handle/10986/6975/350340The0Powe1n0REV01OFFICIAL0USE1.pdf?sequence=1, accessed 03/11/13. Citado aquí

2. McDowell I. Measuring health: a guide to rating scales and questionnaires, 3rd ed. New York: Oxford University Press, 2006. Citado aquí

3. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use, 4th ed. New York: Oxford University Press, 2009. Citado aquí

4. Bland JM, Altman DG. Cronbach's alpha. BMJ 1997;314:572. Citado aquí

5. Mangione CM, Berry S, Spritzer K, et al. Identifying the content area for the 51-item National Eye Institute Visual Function Questionnaire: results from focus groups with visually impaired persons. Arch Ophthalmol 1998;116:227-233. Citado aquí

6. Mangione CM, Lee PP, Pitts J, et al. Psychometric properties of the National Eye Institute Visual Function Questionnaire (NEI-VFQ). NEI-VFQ Field Test Investigators. Arch Ophthalmol 1998;116:1496-1504. Citado aquí

7. Kobak KA, Greist JH, Jefferson JW, et al. Computer assessment of depression and anxiety over the phone using interactive voice response. MD Comput 1999;16:64-68. Citado aquí

8. Dillman DA, Smyth JD, Christian LM. Internet, mail, and mixed-mode surveys: the tailored design method, 3rd ed. Hoboken, NJ: Wiley, 2008. Citado aquí

9. Mackey DC, Manini TM, Schoeller DA, et al. Validation of an armband to measure daily energy expenditure in older adults. J Gerontol A Biol Sci Med Sci 2011;66:1108-1113. Citado aquí

10. Girshik J, Fritschi L, Heyworth J, et al. Validation of self-reported sleep against actigraphy. J Epidemiol 2012; 22:462-468. Citado aquí

CAPÍTULO 16 Gestión de los datos

Michael A. Kohn
Thomas B. Newman
Stephen B. Hulley

Hemos visto que para emprender un proyecto de investigación clínica es necesario elegir un diseño del estudio, definir la población y especificar las variables predictiva y de respuesta. Finalmente, la mayoría de la información sobre los participantes y las variables residirá en una base de datos informática que se usará para almacenar, actualizar y supervisar los datos, así como para dar formato a estos para su análisis estadístico. La base de datos del estudio también puede almacenar datos administrativos, como registros de llamadas, calendarios de visitas y registros de reembolso. Las bases de datos de estudios sencillos, que están formadas por tablas de datos individuales, pueden mantenerse usando hojas de cálculo o programas estadísticos. Las bases de datos más complejas, que contienen múltiples tablas de datos relacionadas entre sí, necesitan un programa de gestión de bases de datos.

La gestión de los datos de un estudio de investigación clínica supone la definición de las tablas de datos, el desarrollo del sistema de introducción de datos y la consulta de estos para su supervisión y análisis. En estudios clínicos extensos, especialmente estudios que se hacen como preparación para la solicitud de autorización por las autoridades de registro sanitario de un fármaco dispositivo, los especialistas que crean los formularios e introducen los datos, gestionan y supervisan el proceso de recogida de datos, y asignan formato y extraen los datos para su análisis se denominan gestores de los datos clínicos (1). Las grandes compañías farmacéuticas que ponen en marcha múltiples ensayos clínicos dedican muchos recursos y personal a la gestión de los datos clínicos. Aunque la escala generalmente es mucho menor, los nuevos investigadores también deben prestar mucha atención a los problemas de la gestión de los datos.

▪ TABLAS DE DATOS

Todas las bases de datos informáticas constan de una o más tablas de datos en las que las filas corresponden a registros (que pueden representar sujetos, episodios o transacciones) y las columnas corresponden a campos (atributos de los registros). Por ejemplo, las bases de datos de los estudios más sencillos constan de una tabla en la que cada fila corresponde a un participante del estudio y cada columna corresponde a un atributo específico del participante, como el nombre, la fecha de nacimiento, el sexo y el estado de la variable predictiva o de respuesta. En general, la primera columna es un número de identificación del participante único («ID partic»). La utilización de un número de identificación único del participante que no tiene ningún significado externo a la base de datos del estudio simplifica el proceso de «desvincular» los datos del estudio de los identificadores personales, para mantener la confidencialidad de los participantes. Si la base de datos contiene tablas adicionales con registros que corresponden a exploraciones, resultados de laboratorio o llamadas telefónicas, entonces la primera columna de cada una de esas tablas debe ser un identificador único del registro, como ID Expl, ID Res Lab o ID Llam. El identificador único del registro de una tabla de datos también se denomina clave principal de la tabla.

La figura 16-1 muestra una tabla de datos simplificada para un hipotético estudio de cohortes (inspirado por un estudio real [2]) de la asociación entre la ictericia neonatal y la puntuación del CI a los 5 años de edad. Cada fila de la tabla corresponde a un participante del estudio, y cada columna a un atributo de este. La variable predictiva dicotómica es si el participante ha tenido o no ictericia («*Jaundice*»), y la variable de respuesta continua es el CI («*IC*»), que es el CI del participante a los 5 años.

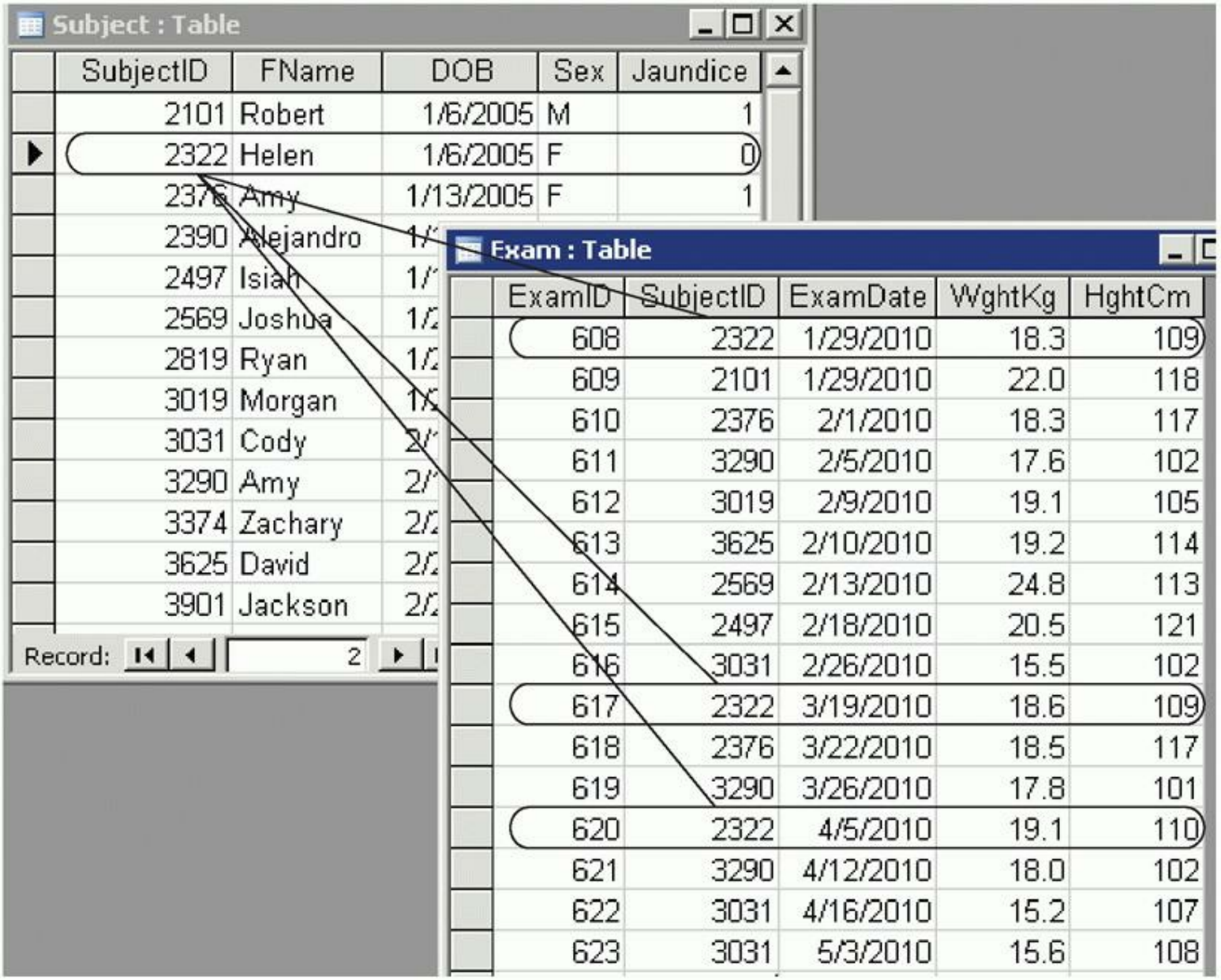
Si los datos del estudio se limitan a una sola tabla como la de la figura 16-1 , se pueden introducir fácilmente en una hoja de cálculo o en un programa de análisis estadístico. A menudo, a una base de datos con una tabla bidimensional se la denomina «archivo plano». Muchos paquetes estadísticos tienen características añadidas para poder utilizar más de una tabla, pero, en su núcleo, la mayor parte siguen siendo bases de datos con archivos planos.

SubjectID	FName	DOB	Sex	Jaundice	ExamDate	WghtKg	HghtCm	IQ
2101	Robert	1/6/2005	M	1	1/29/2010	23.9	118	104
2322	Helen	1/6/2005	F	0	1/29/2010	18.3	109	94
2376	Amy	1/13/2005	F	1	3/22/2010	18.5	117	85
2390	Alejandro	1/14/2005	M	0				
2497	Isiah	1/18/2005	M	0	2/18/2010	20.5	121	74
2569	Joshua	1/23/2005	M	1	2/13/2010	24.8	113	115
2819	Ryan	1/26/2005	M	0				
3019	Morgan	1/29/2005	F	0	2/9/2010	19.1	105	105
3031	Cody	2/15/2005	M	0	4/16/2010	15.2	107	132
3290	Amy	2/16/2005	F	1	4/12/2010	18.0	102	125
3374	Zachary	2/21/2005	M	1				
3625	David	2/22/2005	M	1	2/10/2010	19.2	114	134
3901	Jackson	2/28/2005	M	0				

▪FIGURA 16-1. Tabla de datos simplificada para un estudio de cohortes de la asociación entre ictericia neonatal y puntuación del CI a los 5 años de edad. La variable predictiva dicotómica es «*Jaundice*» («Ictericia»), que se define aquí por un aumento de la bilirrubina total hasta 25 mg/dL o más en los primeros 2 días después del nacimiento, y la variable de respuesta continúa es «IQ», la puntuación del CI a los 5 años de edad. No se evaluó a los 5 años de edad a los participantes 2 390, 2 819, 3 374 y 3 901.
Herramientas de imágenes

La necesidad de incluir más de una tabla en una base de datos de un estudio (y de pasar de la hoja de cálculo o el programa estadístico a un programa de gestión de datos) surge si en el estudio se realiza el seguimiento de múltiples resultados de laboratorio, medicaciones u otras mediciones repetidas en todos los participantes en el estudio. Una única tabla de datos con una fila por participante en el estudio no permite la inclusión de un número grande y variable de mediciones repetidas. La base de datos debe almacenar medicamentos, resultados de laboratorio u otras mediciones repetidas en tablas separadas diferentes de la tabla de los participantes en el estudio. Una fila de una de estas tablas aparte corresponde a una medición individual, como, por ejemplo, el tipo de medición, la fecha/hora de la medición, y el resultado o valor de la medición. Un campo de la fila debe incluir el número de identificación del participante para vincular la medición de nuevo con los campos específicos del participante. En esta «base de datos relacional de múltiples tablas», la relación entre la tabla de participantes y las tablas de las mediciones se denomina de una a muchas. En sentido estricto, el término relacional tiene poco que ver con las relaciones entre las tablas. De hecho, relación es el término formal de la teoría matemática de conjuntos para una tabla de datos (3 , 4).

Aunque a los participantes de nuestro estudio de ictericia en el lactante se les realizó el estudio del CI solo una vez a los 5 años de edad, en la mayoría de ellos se realizaron otras exploraciones, durante las cuales, junto con otras mediciones, se evaluó el peso y la altura. Se utilizaron los datos de la altura y el peso para calcular el índice de masa corporal (IMC) y los percentiles de la curva de crecimiento. (V. «Extracción de datos [consultas]», más adelante en este capítulo.) La mejor forma de poder incluir estos datos es en una tabla separada de exploraciones, en la que cada fila corresponde a una exploración discreta y las columnas representan la fecha de la exploración, los resultados de la exploración y el número de identificación del paciente, para vincularla de nuevo con información de la tabla del participante, como el sexo, la fecha de nacimiento (FN) y si el niño tuvo una ictericia neonatal (fig. 16-2). En esta estructura de bases de datos en dos tablas, la búsqueda en la tabla de exploraciones de todas las exploraciones realizadas en un período temporal particular precisa la búsqueda de una única columna de fecha de exploraciones. En un lugar se hace un cambio de un campo específico del participante, como la fecha de nacimiento, y se mantiene la congruencia. Los campos que contienen identificadores personales, como el nombre y la fecha de nacimiento, aparecen únicamente en la tabla del participante. Las otras tablas establecen un vínculo con esta información gracias al número de identificación del participante. La base de datos permite la introducción de participantes (como Alejandro, Ryan, Zachary y Jackson) en los que no se han realizado exploraciones.



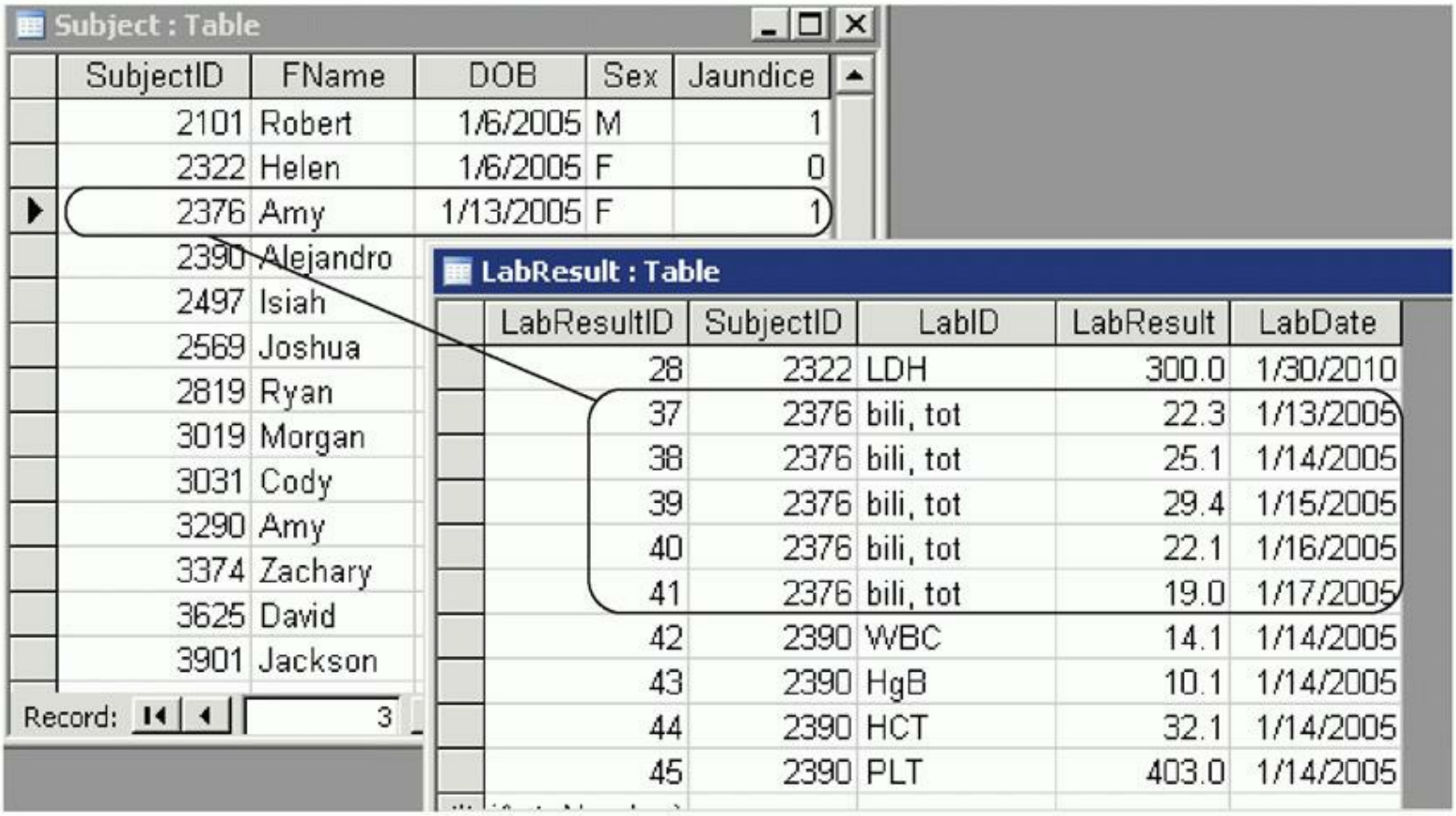
▪FIGURA 16-2. La base de datos de dos tablas del estudio de la ictericia neonatal tiene una tabla de participantes del estudio, en la que cada fila corresponde a un solo participante del estudio, y una tabla de exploraciones, en la que cada fila corresponde a una exploración concreta. Por ejemplo, se identifica a la participante 2 322 como Helen, con fecha de nacimiento el 6 de enero de 2005, en la primera tabla, y se muestran datos de las tres evaluaciones en la segunda tabla anónima. Como pueden haberse realizado múltiples exploraciones a un participante, la relación entre las dos tablas es de una a muchas. El campo «*SubjectID*» («Identidad del participante») en la tabla de evaluaciones vincula los datos específicos de la evaluación con los datos específicos del participante.
Herramientas de imágenes

El seguimiento detallado de los resultados de laboratorio también precisa una tabla separada. Aquí se presenta la ictericia neonatal como un campo dicotómico específico de participante. Si los investigadores necesitan toda la trayectoria de la concentración de la bilirrubina después del nacimiento, entonces la base de datos debe incluir una tabla de resultados de laboratorio separada con un registro para cada resultado de laboratorios y campos para la fecha/hora de la prueba de laboratorio, el tipo de prueba de laboratorio (bilirrubina total), el resultado de la prueba (concentración de bilirrubina) e identificación del participante para relacionar esta información con la información específica del participante (fig. 16-3).

Los datos administrativos del estudio, como registros de llamadas, calendarios de visitas y registros de reembolso, también precisan múltiples tablas separadas. En el estudio de ictericia en lactantes se hicieron múltiples llamadas a los padres de cada uno de los participantes en el estudio. Sería difícil o imposible hacer un seguimiento de estas llamadas en una tabla de datos con una fila para cada participante del estudio. Por el contrario, en una tabla separada había una fila para cada llamada, con un campo de identificación del participante que relacionaba esta tabla con el participante en el estudio sobre el cual se hacía la llamada.

La estructuración de la base de datos con múltiples tablas relacionadas, en lugar de intentar introducir los datos en una tabla única muy ancha y compleja, se denomina normalización. Algunos gestores de datos se refieren a la normalización como la conversión de una o

unas pocas tablas «bajas y gordas» en muchas tablas «altas y delgadas» (1). La normalización elimina el almacenamiento redundante y la posibilidad de incongruencias. Los programas de bases de datos relacionales mantendrán la integridad referencial, lo que significa que no se permitirá la creación de un registro de una exploración, resultado de laboratorio o registro de llamadas de un participante que no exista todavía en la tabla de participantes. Igualmente, evita la eliminación de un participante, salvo que también se hayan eliminado todas las exploraciones, resultados de laboratorio y llamadas de ese participante.

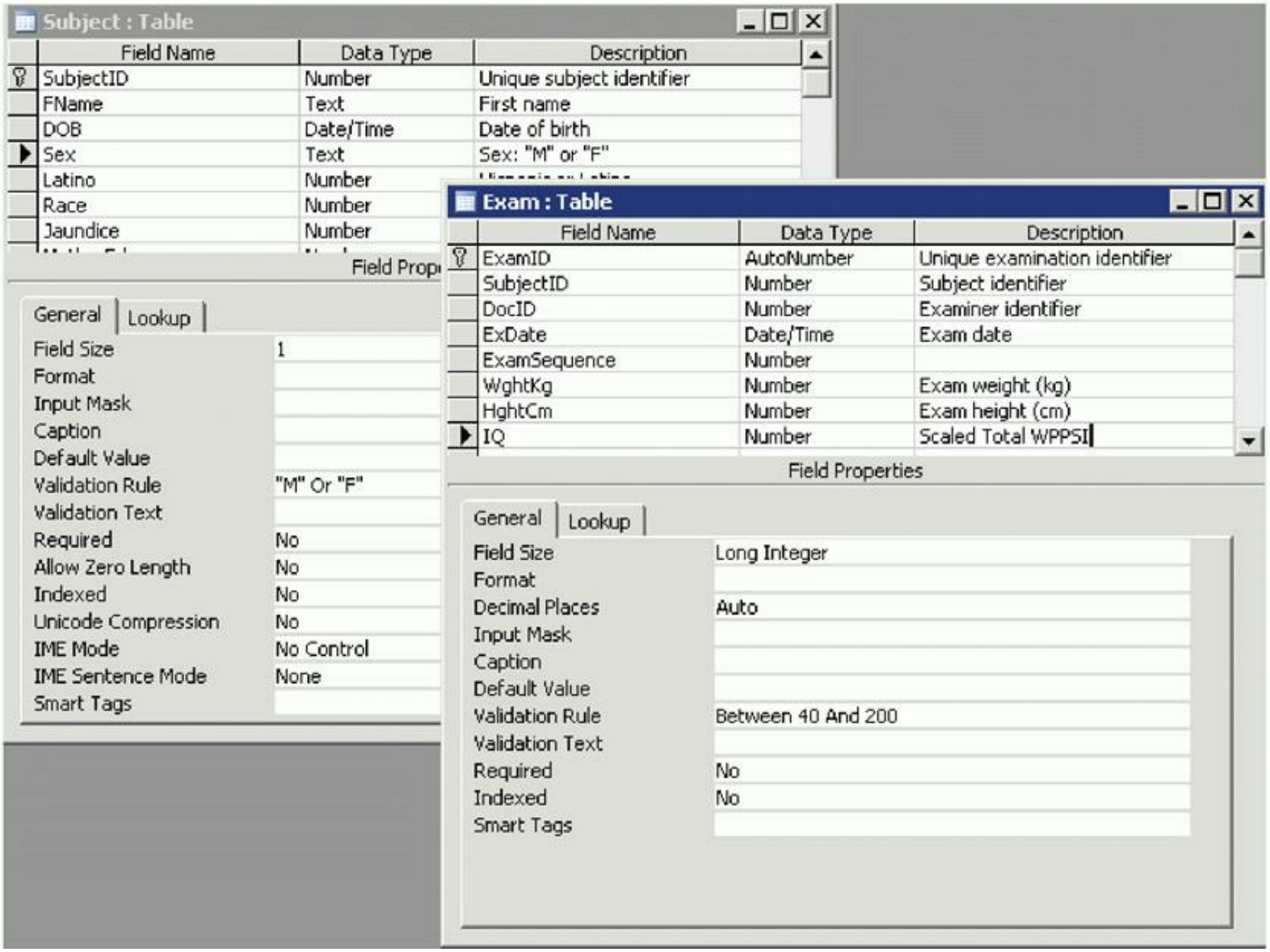


▪FIGURA 16-3. Relación entre la tabla de participantes y la tabla de resultados de laboratorio. Los resultados de laboratorio permiten ver la trayectoria de la bilirrubina total de Amy en los primeros 5 días después del nacimiento.
Herramientas de imágenes

Diccionarios de datos, tipos de datos y dominios

Hasta ahora, hemos visto tablas solo en la vista de «hoja de cálculo». Cada columna o campo tiene un nombre e, implícitamente, un tipo de datos y una definición. En la tabla «Participante» («Subject») de la figura 16-2 , «FName» es un campo de texto que contiene el nombre del participante; «DOB» (FN) es un campo de fecha que contiene la fecha de nacimiento del participante, y «*Jaundice*» («Ictericia») es un campo de sí/no que indica si la bilirrubina fue mayor de 25 mg/dL en los primeros 2 días después del nacimiento. En la tabla «*Exam*» («Exploración»), «*WghtKg*» es el peso en números reales en kilogramos, e «*IQ*» (CI) es un entero que corresponde a la puntuación del CI. El diccionario de los datos explicita las definiciones de estas columnas. En la figura 16-4 se muestran las tablas de participantes y de exploraciones en vista de diseño de tabla (o «diccionario de los datos»). Observe que el diccionario de los datos en sí mismo es una tabla con filas, que representan campos, y columnas, que representan el nombre, el tipo y la descripción de los campos. Ya que el diccionario de los datos es una tabla de información sobre la propia base de datos, se denomina metadatos. Aunque en la figura 16-4 se ven dos diccionarios de datos, uno para la tabla «*Subject*» («Participante») y otro para la tabla «*Exam*» («Exploraciones»), se puede considerar que toda la base de datos tiene un único diccionario de los datos, en lugar de un diccionario para cada tabla. Para cada uno de los campos de la base de datos, el diccionario de los datos único precisa que se especifique el nombre de la tabla del campo, además del nombre del campo, el tipo de campo, la descripción del campo y el intervalo de valores permitidos.

Cada campo tiene también un dominio o intervalo de valores permitidos. Por ejemplo, los valores permitidos para el campo «*Sex*» («Sexo») son «*M*» («V») y «*F*» («M»). El programa informático no permitirá la entrada de ningún otro valor en este campo. Igualmente, el campo «*IQ*» solo permite enteros entre 40 y 200. Los gestores de los datos de ensayos clínicos generalmente se refieren a las reglas de validación como «verificaciones de la edición» (1). La creación de normas de validación para definir los valores permitidos proporciona cierta protección frente a los errores en la introducción de los datos. Algunos de los tipos de datos vienen con normas de validación automáticas. Por ejemplo, los programas informáticos para la gestión de bases de datos siempre rechazará una fecha de 31 de abril.



▪FIGURA 16-4. La tabla de los participantes en el estudio («*Subject*») («Participante») y la tabla de las mediciones («*Exam*») («Exploración») en la vista del «diccionario de los datos». Cada variable o campo tiene un nombre, un tipo de dato, una descripción y un dominio o conjunto de valores permitidos.
Herramientas de imágenes

Nombres de variables

La mayor parte de las hojas de cálculo, y de los programas estadísticos y de gestión de bases de datos, permiten encabezamientos de columnas o nombres de variables largos. Abundan las filosofías y las convenciones para la asignación de nombres. Recomendamos nombres de variables lo suficientemente cortos para poder escribirlos rápidamente, pero lo suficientemente largos como para que sean autoexplicativos. Aunque muchas veces lo permite el programa informático, recomendamos evitar los espacios y los caracteres especiales en los nombres de las variables. Distinguimos las palabras separadas en el nombre de una variable utilizando lo que acertadamente se denomina «*InterCaps*» («Mayúsculas internas»), aunque otros autores prefieren utilizar un guion bajo. Suele ser mejor usar un nombre de variable que describa el campo, en lugar de su ubicación en el formulario de recopilación de datos (p. ej., «AlgunaVezFumaCigarrillos», en lugar de «Pregunta1»). La mayoría de los programas informáticos permiten que los usuarios utilicen una etiqueta de variable más larga, descriptiva y fácil de leer para utilizar en los formularios de introducción de datos y los informes, en lugar del nombre de la variable compacto.

Elementos de datos comunes

Algunas organizaciones que financian ensayos clínicos y agencias de registro sanitario han puesto en marcha iniciativas para elaborar elementos de datos comunes para bases de datos de estudios en áreas específicas de la investigación clínica. Entre estas organizaciones están agencias gubernamentales, como el National Institute for Neurologic Disorders and Stroke (5), el National Cancer Institute (6), la United States Food and Drug Administration (7) y la Agencia Europea del Medicamento, y asociaciones no gubernamentales sin ánimo de lucro, como el Clinical Data Interchange Standards Consortium (CDISC) (8).

La justificación es que en los estudios de investigación en una misma área clínica muchas veces se deben obtener las mismas mediciones. La normalización de las estructuras de los registros, los nombres y las definiciones de campos, los tipos y formatos de datos, y los formularios de recogida de datos (cuadernos de recogida de datos) elimina el problema de «reinventar la rueda», como muchas veces ocurre en nuevos estudios de investigación (5), y permite que se compartan y combinen datos de múltiples estudios separados. Esto incluye el establecimiento de un diccionario de los datos y un conjunto de instrumentos para la recogida de datos con las correspondientes instrucciones, que se anima a que utilicen todos los investigadores en un área de investigación particular. Una parte de la erudición profunda en el área de investigación elegida es el conocimiento de las normas existentes en relación con los datos.

Volver al principio

▪ INTRODUCCIÓN DE DATOS

Tanto si la base de datos del estudio consta de una como de muchas tablas, y tanto si se usa una hoja de cálculo, un programa estadístico o un programa de gestión de bases de datos, se necesita un mecanismo para rellenar las tablas de datos (introducir los datos).

Transcripción con el teclado

Históricamente, el método habitual para rellenar una base de datos de un estudio ha sido recopilar datos en formularios de papel. En los ensayos clínicos, el formulario para recogida de datos en papel que corresponde a un participante específico se denomina cuaderno de recogida de datos o CRD. El investigador y otros miembros del equipo de investigación pueden rellenar el formulario en papel o, en algunos casos, lo hace el propio participante. Después, el personal del estudio puede transcribir los datos con el teclado desde los formularios en papel hasta las tablas del ordenador. La transcripción puede realizarse directamente a las tablas de datos (p. ej., la respuesta a la pregunta 3 en el participante 10 va en la celda de la fila 10, columna 3) o mediante formularios en pantalla, diseñados para facilitar la entrada de datos y que incluyen comprobaciones de validación de datos automáticas. La transcripción debe realizarse lo más pronto posible tras la recopilación de datos, de modo que el participante y el entrevistador o recopilador de los datos sigan estando disponibles si se observa que las respuestas se pierden o están fuera del intervalo de valores. También, como se comenta más adelante en este capítulo, solo puede realizarse la supervisión de los problemas con los datos (p. ej., valores atípicos) y los análisis preliminares una vez que los datos estén en la base de datos del ordenador.

Si se transcribe desde formularios en papel, el investigador debe considerar la doble introducción de datos para asegurar la fidelidad de la transcripción. El programa de bases de datos compara los dos valores introducidos para cada variable y presenta una lista de los valores que no encajan. Las entradas discrepantes se comprueban en los formularios originales y se corrigen. La doble entrada de datos identifica errores en la introducción de los datos a costa de duplicar el tiempo necesario para su transcripción. Una alternativa es la doble introducción de los datos de una muestra aleatoria. Si la tasa de error es aceptablemente baja, es poco probable que esté justificado el esfuerzo y el coste de la doble introducción de los datos restantes.

Entrada distribuida de datos

Si la recogida de datos se realiza en múltiples centros, los centros pueden enviar por correo electrónico o por fax los formularios en papel a una oficina centralizada para su transcripción a la base de datos informática, aunque esta práctica es cada vez menos frecuente. La mayoría de las veces, los datos se transcriben directamente en la base de datos del estudio en los propios centros de investigación con formularios en Internet. Si hay problemas en la conexión a Internet, los datos se almacenan en un ordenador local en el centro y se transmiten por vía electrónica o con un dispositivo de memoria portátil, como una memoria USB. Las regulaciones gubernamentales exigen que la información sanitaria electrónica vaya desprovista de identificación o que se transmita con medidas de seguridad (p. ej., encriptada y protegida con una contraseña).

Captura de datos electrónicos

La recogida de los actos principales en papel siempre tendrá su lugar en la investigación clínica; una forma rápida y amigable de recoger datos en un medio no volátil es utilizar pluma y papel. Sin embargo, escribir a mano los datos en un formulario en papel es cada vez menos frecuente. En general, en los estudios de investigación se recopilan datos usando formularios en pantalla. En los ensayos clínicos, los formularios electrónicos se denominan cuadernos de recogida de datos electrónicos (CRDe). La recogida de los datos con formularios en pantalla tiene muchas ventajas:

- Los datos se teclean directamente en tablas de datos sin un segundo paso de transcripción, lo que elimina esa fuente de error.
- El formulario informático puede incluir comprobaciones de validación y proporcionar una retroalimentación inmediata cuando se introduce un dato que está fuera del intervalo de valores.

- El formulario informático también puede incorporar lógica de exclusión. Por ejemplo, una pregunta sobre el número de cajetillas al día aparecerá solo si el participante responde «sí» a una pregunta sobre consumo de cigarrillos.

- El formulario puede verse y se pueden introducir los datos en aparatos portátiles e inalámbricos, como una tableta (iPad), un teléfono inteligente o un ordenador portátil.

Al usar formularios en pantalla para recoger datos electrónicos, a veces es útil hacer una copia en papel de los datos inmediatamente tras la recopilación. Es lo mismo que imprimir un recibo tras una transacción en un cajero automático. La impresión es una «instantánea» en papel del registro inmediatamente después de recopilar los datos, y puede usarse como documento original si se precisa una versión en papel.

Respuestas codificadas y texto libre

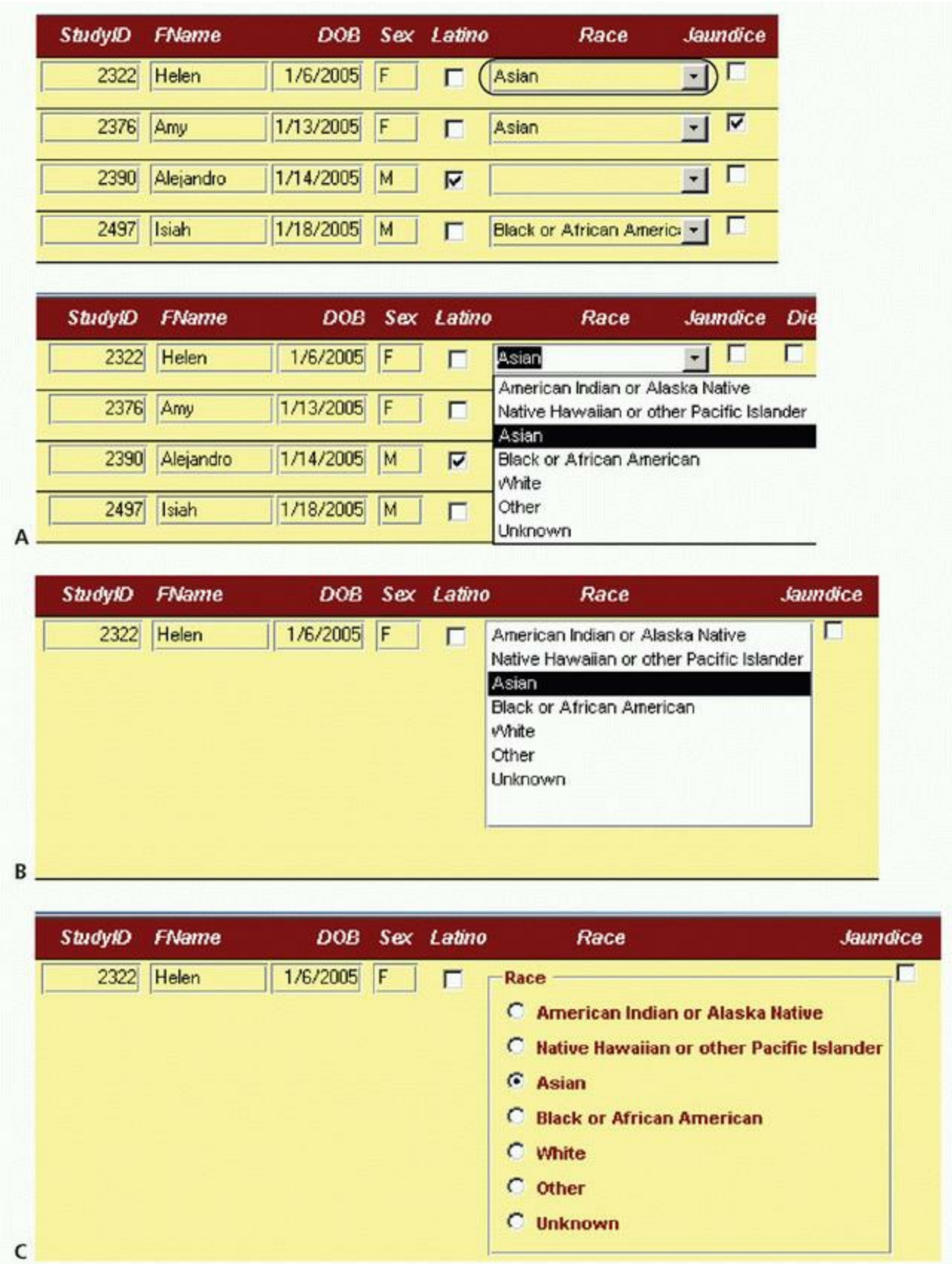
Definir una variable o campo en una tabla de datos conlleva la especificación de su intervalo de valores permitidos. Para posteriores análisis es preferible limitar las respuestas a un intervalo de valores codificados, en lugar de permitir respuestas de texto libre. Es lo mismo que la distinción que se hace en el capítulo 15 entre preguntas «abiertas» y «cerradas». Si no está claro el intervalo de posibles respuestas, la recopilación inicial de datos durante la comprobación previa del estudio puede permitir respuestas de texto libre, que se usarán, posteriormente, para desarrollar opciones de respuesta codificadas.

La serie de opciones de respuesta a una pregunta debe ser exhaustiva (se proporcionan todas las opciones posibles) y mutuamente excluyente (no hay dos opciones que puedan ser a la vez correctas). Siempre se puede conseguir que un conjunto de opciones de respuesta mutuamente excluyentes sean exhaustivas en forma colectiva añadiendo una respuesta de «otros». Los formularios de recogida de datos en pantalla proporcionan tres posibles formatos para distribuir las opciones de respuesta colectivamente exhaustivas y mutuamente excluyentes: lista desplegable, lista de selección o grupo de opciones (fig. 16-5). Estos formatos serán familiares para cualquier participante en una investigación o cualquier persona que introduzca datos y que haya trabajado con formularios informáticos. Debe tenerse en cuenta que la lista desplegable ahorra espacio de pantalla, pero no funcionará si el formulario de pantalla tiene que imprimirse en papel para la recopilación de datos, porque no se podrán ver las opciones de respuesta.

Una pregunta con una serie de respuestas mutuamente excluyentes corresponde a un solo campo en la tabla de datos. Por el contrario, las respuestas a una pregunta de «Todas las que corresponda» no son mutuamente excluyentes. Corresponden a tantos campos de sí/no como respuestas posibles haya. Por convención, las opciones de respuesta de «Todas las que corresponda» usan casillas de verificación en lugar de los botones redondos que se usan para los grupos de opciones con respuestas mutuamente excluyentes. Como ya se ha señalado en el capítulo 15, desaconsejamos las preguntas de «Todas las que corresponda», y preferimos que haya que introducir una respuesta de sí o no a cada pregunta. En caso contrario, una respuesta no marcada podría significar «no procede» o «no respondida». Cuando se codifican las variables de tipo sí/no (dicotómica), hay que hacer que 0 represente *no* o *ausente*, y 1 represente *sí* o *presente*. Con esta codificación se puede interpretar el valor medio de la variable como la proporción con el atributo.

Importación de mediciones y resultados de laboratorio

Gran parte de la información del estudio, como la información demográfica basal en el sistema de registro del hospital, los resultados de laboratorio en el sistema informático del laboratorio y las determinaciones realizadas mediante absorciometría de rayos X de doble energía (DEXA) y los registros Holter, ya se encuentra en formato electrónico digital. Cuando sea posible, estos datos deben importarse directamente a la base de datos del estudio, para evitar el trabajo y los posibles errores de transcripción que conlleva volver a introducir los datos. Por ejemplo, en el estudio de la ictericia neonatal, los datos demográficos de los participantes y la información de contacto se obtuvieron de la base de datos del hospital. Los sistemas informáticos casi siempre pueden producir archivos de texto delimitados por tabuladores o con anchura de columna fija que pueden importar los programas de bases de datos. En los ensayos clínicos, este tipo de información actualizada por lotes se denomina «datos no obtenidos con CRD (cuaderno de recogida de datos)» (1).



▪FIGURA 16-5. Formatos para la introducción en una lista de respuestas mutuamente excluyentes y exhaustivas de manera colectiva. La lista desplegable (A; desplegada en el panel inferior) ahorra espacio en la pantalla, aunque no funcionará si el formulario de pantalla se va a imprimir para la recogida de datos. Hace falta más espacio tanto para la lista de selección (que es tan solo una lista desplegable que está desplegada continuamente; B) como para el grupo de opciones (C), aunque sí se pueden utilizar cuando se imprimen.

Herramientas de imágenes

Hojas de cálculo

Microsoft Excel

Google Drive Spreadsheet*

Apache OpenOffice Calc*

Análisis estadístico

Statistical Analysis System (SAS)

Statistical Package for the Social Sciences (SPSS)

Stata

R*

EpiInfo* (solo para Windows)

Sistemas de bases de datos de escritorio integradas

Microsoft Access (solo para Windows)

Filemakers Pro

Sistemas de bases de datos relacionales

Oracle

SQL Server

MySQL*

PostgreSQL*

Aplicaciones integradas basadas en la web para el manejo de datos de investigación

Research Electronic Data Capture* (REDCap: solo para uso académico, alojado por la institución del investigador)

QuesGen (principalmente académico, alojado en el proveedor)

MediData RAVE (principalmente corporativo no académico, alojado en el proveedor)

Oracle InForm (corporativo no académico, alojado en la compañía)

Datalabs EDC (corporativo, alojado en el proveedor)

OnCore

OpenClinica

Herramientas para realizar encuestas en Internet

SurveyMonkey

Zoomerang

Qualtrics

* Gratuitos

Programas informáticos para la gestión de datos

Una vez comentadas las tablas de datos y la introducción de datos, se puede hacer la distinción entre los sistemas de soporte (*back-end*) y frontal (*front-end*) de las bases de datos del estudio. El sistema de soporte consiste en las propias tablas de datos. El sistema frontal o «interfaz» está formado por los formularios informáticos que se usan para introducir, ver y editar los datos. En la tabla 16-1 se ofrece una lista de algunas aplicaciones informáticas que se usan para la gestión de datos en investigación clínica. Las bases de datos de estudios sencillos constan de una sola tabla de datos que puede usar una hoja de cálculo o un programa estadístico para la tabla de datos de soporte, y el personal del estudio puede introducir datos directamente en las celdas de la tabla, evitando la necesidad de los formularios de recopilación de datos del sistema frontal. Las bases de datos de estudios más complejos constan de múltiples tablas de datos que necesitan un programa de bases de datosrelacionales para mantener las tablas de datos de soporte. Si los datos se recogen primero en formularios de papel, la introducción de los datos precisará la transcripción en formularios informáticos.

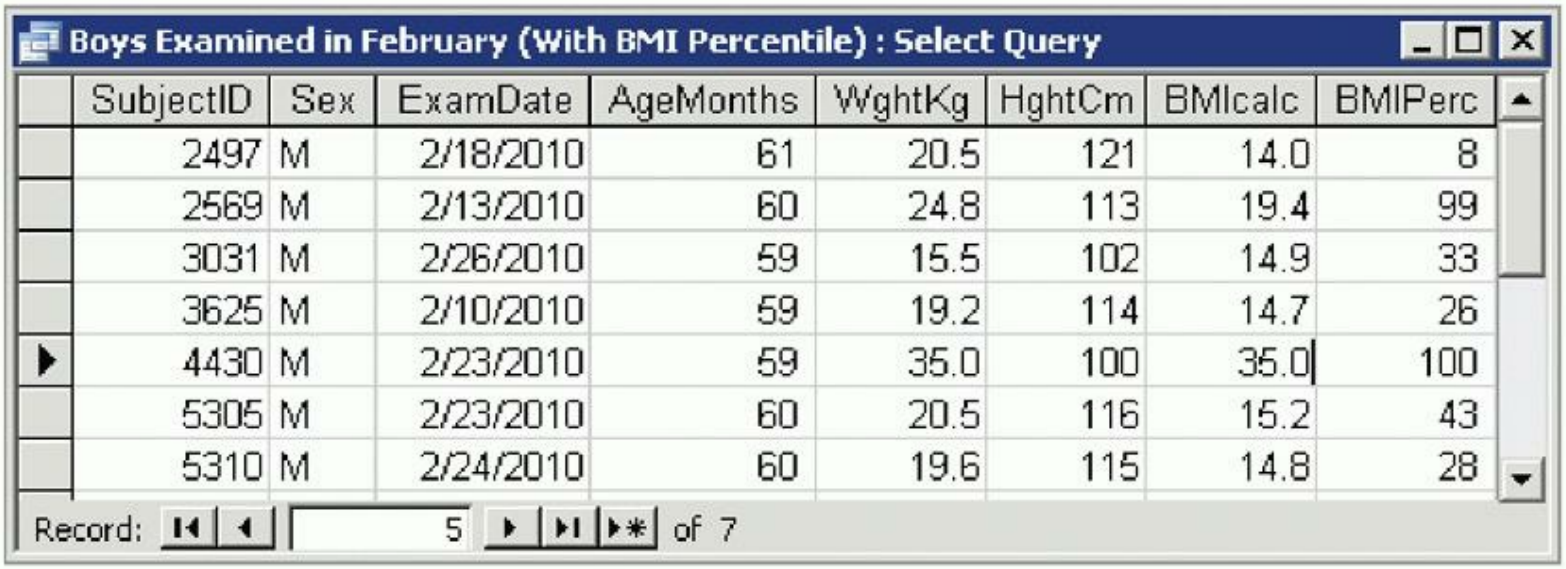
Como se comentó en el capítulo 15, existen varias herramientas, como SurveyMonkey, Zoomerang y Qualtrics, para desarrollar encuestas en línea que se enviarán por correo electrónico a los participantes en el estudio o se cargarán en la página web del estudio. Todas estas herramientas ofrecen opciones de formato de preguntas múltiples, lógica de exclusión y la posibilidad de agregar, describir y exportar los resultados de la encuesta.

Para algunos de los paquetes estadísticos, como SAS, se han desarrollado módulos para la introducción de los datos. Los programas de bases de datos de escritorio integradas, como Microsoft Access y Filemarker Pro, también proporcionan muchas herramientas para el desarrollo de formularios de datos.

En los estudios de investigación cada vez se utilizan más plataformas de gestión de datos de investigación integradas y con posibilidades de uso en Internet. REDCap (*Research Electronic Data Capture*) es un sistema de recogida de datos de investigación basado en Internet y desarrollado por un consorcio académico de la Universidad de Vanderbilt. Permite que los investigadores elaboren formularios de introducción de datos, encuestas y encuestas con formularios adjuntos para la introducción de datos. REDCap está disponible únicamente para investigadores académicos y se debe alojar en el centro del investigador. Es una notable herramienta casera para nuevos investigadores académicos que permite el desarrollo rápido de encuestas y formularios de recogida de datos en pantalla. También permite acceder a un almacén de instrumentos de recogida de datos que se pueden descargar. Como todas las herramientas caseras de desarrollo en Internet, las opciones de personalización y funcionalidad avanzada son escasas. Una base de datos elaborada con REDCap

está formada por una única tabla que contiene una hilera para cada uno de un número fijo de «acontecimientos» definidos por el usuario para cada uno de los participantes en el estudio. No permite un seguimiento detallado de un número grande y variable de mediciones repetidas de cada uno de los participantes en el estudio, como resultados de laboratorio, constantes vitales, medicamentos o registros de llamadas. Con el sistema REDCap tampoco se pueden realizar procedimientos sofisticados de validación, consulta (v. más adelante en este capítulo) o informes de los datos, aunque facilita la exportación a paquetes estadísticos.

Las plataformas de gestión de datos de investigación completas y basadas en Internet, como QuesGen, MediData RAVE y Oracle InForm, permiten la introducción de estructuras de datos complejos y ofrecen herramientas sofisticadas para la validación, la consulta y el informe de los datos. Las compañías que suministran estas herramientas también ofrecen soporte técnico y ayuda para la configuración. Aunque puede suponer algún coste adicional, merece la pena tener en consideración estas soluciones cuando las herramientas caseras carecen de la sofisticación que permite satisfacer las necesidades del estudio.



Boys Examined in February (With BMI Percentile) : Select Query

SubjectID	Sex	ExamDate	AgeMonths	WghtKg	HghtCm	BMIcalc	BMIPerc
2497	M	2/18/2010	61	20.5	121	14.0	8
2569	M	2/13/2010	60	24.8	113	19.4	99
3031	M	2/26/2010	59	15.5	102	14.9	33
3625	M	2/10/2010	59	19.2	114	14.7	26
4430	M	2/23/2010	59	35.0	100	35.0	100
5305	M	2/23/2010	60	20.5	116	15.2	43
5310	M	2/24/2010	60	19.6	115	14.8	28

Record: 5 of 7

▪FIGURA 16-6. Una consulta en vista de hoja de cálculo que filtra los resultados correspondientes a los niños explorados en febrero y calcula la edad en meses (desde la fecha de nacimiento hasta la fecha de la exploración), además del índice de masa corporal (*BMI*; IMC) calculado a partir del peso y la altura. La consulta también utiliza una sofisticada función de búsqueda en la tabla para calcular los valores del percentil de la curva de crecimiento correspondientes al IMC del niño. Para el paciente con el número de identificación 4 430, el valor del percentil 100 asociado al IMC de 35,0 debería llevar a investigar el valor extremo como un posible error en la introducción de los datos.
Herramientas de imágenes

Volver al principio

▪ EXTRACCIÓN DE DATOS (CONSULTAS)

Una vez que se ha creado la base de datos y que se han introducido estos, el investigador deseará organizar, clasificar, filtrar y ver («consultar») los datos. Las consultas se utilizan para supervisar la introducción de los datos, comunicar el progreso del estudio y, finalmente, analizar los resultados. El lenguaje habitual para manipular datos en una base de datos relacional se denomina lenguaje de consulta estructurado (SQL,*Structured Query Language*). Todos los sistemas informáticos de bases de datos relacionales usan una u otra variante de SQL, pero la mayor parte disponen de una interfaz gráfica para construir consultas que hace que no sea necesario que los investigadores clínicos aprendan SQL.

Una consulta puede combinar datos de dos o más tablas, mostrar solo campos seleccionados y filtrar los registros que cumplen ciertos criterios. Las consultas también pueden calcular valores basados en los campos de datos brutos de las tablas. La figura 16-6 muestra los resultados de una consulta en nuestra base de datos de ictericia neonatal que filtra los niños varones estudiados en febrero, y calcula la edad en meses (desde la fecha de nacimiento y la fecha del examen) y el IMC (a partir del peso y la altura). La consulta también utiliza una función sofisticada de búsqueda en las tablas para calcular los valores de los percentiles de las curvas de crecimiento para el IMC del niño. Observe que el resultado de una consulta que une dos tablas, proporciona solo determinados campos, selecciona filas según criterios especiales y calcula determinados valores, sigue pareciendo una tabla en la vista de hoja de cálculo. Uno de los principios del modelo de bases de datos relacionales es que las operaciones en las tablas producen resultados tabulares. Los datos de la figura 16-6 se exportan fácilmente a un paquete de análisis estadístico. Observe que no se incluyen identificadores personales en la consulta.

Identificación y corrección de errores en los datos

El primer paso para evitar los errores en los datos es verificar su recogida y el sistema de gestión como parte de la comprobación general previa al estudio. Debe estudiarse todo el sistema (tablas de datos, formularios para la introducción de datos y consultas) usando datos simulados. Para los ensayos clínicos que se utilizarán para una solicitud a la FDA, se trata de un requisito de las agencias de registro sometido al código de regulaciones federales (*Code of Federal Regulations*), capítulo 21, parte 11 (21 CFR 11) (9).

Ya se han comentado métodos para aumentar la fidelidad de la transcripción con teclado o con captura electrónica de los datos una vez que ha comenzado su recogida. Los valores que se encuentran fuera de los valores permisibles no deben ir más allá del proceso de entrada de datos. Sin embargo, también deben hacerse búsquedas en la base de datos por si hay valores ausentes y atípicos (valores extremos que se encuentran, no obstante, dentro de los límites de los valores permitidos). Por ejemplo, un peso de 35 kg podría estar dentro de los valores permitidos para un niño de 5 años, pero si es 5 kg superior a cualquier otro peso en la serie de datos, debe llevar a una investigación. Muchos sistemas de introducción de datos no permiten realizar una validación cruzada, lo que significa que las tablas de datos pueden contener valores de campos que se encuentran dentro de los límites permitidos, pero que no con coherentes entre sí. Por ejemplo, no tendría sentido que un niño de 5 años de edad y de 35 kg de peso tuviera una altura de 100 cm. Aunque los valores de peso y altura están dentro de los límites permitidos, el peso (enormemente elevado para 5 años) no es coherente con la altura (enormemente baja para un niño de 5 años). Se puede sospechar esta incongruencia utilizando una consulta como la que se muestra en la figura 16-6 .

Los valores ausentes, los valores atípicos, las incoherencias y otros problemas de los datos se identifican con las consultas y se comunican al personal del estudio, que puede responder a ellos comprobando los documentos originales, entrevistando al participante o repitiendo la determinación. Si el estudio se encuentra en documentos en papel, cualquier cambio de los datos deberá resaltarse (p. ej., en rojo), fecharse y firmarse. Como se comenta más adelante en este capítulo, en las bases de datos electrónicas se debe mantener un registro de auditoría de todos los cambios de los datos.

Si los datos son recopilados por varios investigadores de diferentes consultas o localizaciones, deben compararse las medias y las medianas entre investigadores y lugares. Si se observan diferencias importantes entre investigadores o entre localizaciones, puede indicar que existen diferencias sistemáticas en cuanto a la medición o la recopilación de los datos.

La edición y la limpieza de los datos deben dar mayor prioridad a las variables más importantes. Por ejemplo, en un ensayo aleatorizado, la variable más importante es la de respuesta, por lo que se deben reducir al mínimo los datos ausentes y los errores. Por el contrario, los errores en otras variables, como la fecha de una consulta, pueden no afectar mucho a los resultados de los análisis. La edición de los datos es un proceso iterativo; tras identificar y corregir los errores, los procedimientos de edición deben repetirse hasta que se identifiquen errores muy poco importantes. En este punto, se considera que la base de datos editada es la versión final o que está «cerrada», de modo que ya no se permiten más cambios (1).

Volver al principio

▪ ANÁLISIS DE LOS DATOS

El análisis de los datos a menudo necesita la creación de nuevas variables derivadas basadas en los valores brutos de los campos de la serie de datos. Por ejemplo, las variables continuas pueden dicotomizarse (p. ej., IMC > 25 se define como sobrepeso), pueden crearse nuevas categorías (fármacos específicos agrupados como antibióticos) y pueden realizarse cálculos (años de tabaquismo × número de cajetillas de cigarrillos al día = cajetillas/años). Los datos ausentes se deben manejar de manera constante. «No lo sabe» se puede registrar como una categoría especial, combinado con «no», o excluido como dato ausente. Si el estudio usa un programa de bases de datos, pueden usarse consultas para derivar las nuevas variables antes de exportarlas a un paquete de análisis estadístico. Esto es especialmente importante para variables como los percentiles de la figura 16-6 , en los que hace falta una programación compleja o una tabla de «búsqueda» separada. Otra solución es realizar la derivación de los nuevos campos con el propio paquete estadístico. Muchos investigadores están más familiarizados con los paquetes estadísticos que con los programas de bases de datos, y prefieren calcular las variables derivadas después de la exportación de los datos.

Volver al principio

▪ CONFIDENCIALIDAD Y SEGURIDAD

Si los participantes en un estudio de investigación también son pacientes de una consulta o un hospital, la información que les identifica está protegida por la norma de privacidad (*Privacy Rule*) de la ley de portabilidad y responsabilidad de los seguros sanitarios (*Health Insurance Portability and Accountability Act*, HIPAA) (10); una vez dicho esto, independientemente de que los participantes

también sean pacientes, el investigador tiene la obligación ética y legal de proteger su confidencialidad. La base de datos debe asignar a cada uno de los participantes un identificador de participante único (identidad del participante) que no tenga ningún significado externo a la base de datos del estudio (es decir, la identidad del participante no debe incorporar el nombre, las iniciales, la fecha de nacimiento ni el número de historia clínica del participante). Todos los campos de la base de datos que contengan identificadores personales se deben borrar antes de compartir los datos. Si la base de datos utiliza múltiples tablas, los identificadores personales se pueden conservar en una tabla separada. Las bases de datos del estudio que contengan identificadores personales se deben mantener en servidores seguros a los que puedan acceder únicamente miembros autorizados del equipo de investigación, cada uno de los cuales tendrá una identificación de usuario y una contraseña. Las plataformas específicas para la gestión de datos de la investigación basadas en Internet, como REDCap y QuesGen, permiten la designación de campos que contengan identificadores de los participantes. Diferentes tipos de usuario pueden permitir o prohibir la exportación, la modificación o incluso la visualización de estos campos con designación especial.

El sistema de bases de datos debe realizar una auditoría de toda la introducción y edición de los datos. La auditoría permite la determinación de cuándo se ha modificado un elemento de datos, quién realizó el cambio y qué cambios se realizaron. Para los ensayos de nuevos fármacos, se trata de un requisito de las autoridades de registro (9). Las plataformas de investigación específicas basadas en Internet, como REDCap, QuesGen y MediData RAVE, ofrecen automáticamente la validación y la auditoría del usuario.

La base de datos del estudio debe guardarse frecuentemente y depositarse en otro lugar. Periódicamente, el procedimiento de copia de seguridad debe comprobarse restableciendo una copia de seguridad de los datos. Como con la validación y la auditoría del usuario, las plataformas alojadas en servidores externos, como REDCap, QuesGen y MediData RAVE, ofrecen automáticamente copias de seguridad y sistemas de seguridad de los datos. Al final del estudio se deben archivar para su uso futuro los datos originales, el diccionario de los datos, la base de datos final y los análisis del estudio.

Estos archivos pueden revisarse en los años siguientes, lo que permitirá al investigador responder a preguntas sobre la integridad de los datos o el análisis, realizar otros análisis para abordar nuevas preguntas de investigación y compartir datos con otros investigadores.

Volver al principio

▪ RESUMEN

- La base de datos del estudio consta de una o más tablas de datos en las que las filas corresponden a registros (es decir, participantes en el estudio) y las columnas corresponden a campos (atributos de los registros).
- La identificación de los participantes en el estudio con una identificación del participante única que no tenga ningún significado externo a la base de datos del estudio permite la «desvinculación» de los datos el estudio respecto a los identificadores personales para mantener la confidencialidad. Las bases de datos que contienen identificadores personales se deben almacenar en servidores seguros, con acceso restringido y sometidos a auditorías.
- La introducción de un número variable de medidas repetidas para cada uno de los participantes en el estudio, como resultados de laboratorio o medicamentos, precisa la normalización de los datos de las mediciones en tablas separadas en las que cada fila corresponde a una medición, no a un participante del estudio individual.
- La base de datos del estudio también puede almacenar datos administrativos, como registros de llamadas, calendarios de exploraciones y registros de reembolso.
- El diccionario de los datos especifica el nombre, el tipo de datos, la descripción y el intervalo de valores permitidos para todos los campos de la base de datos.
- El sistema de introducción de los datos es el medio por el que se rellenan las tablas de datos; la captura electrónica de los datos con formularios en pantalla está sustituyendo a la transcripción de esos formularios en papel para la introducción de los datos.
- Una hoja de cálculo o un paquete estadístico es adecuado únicamente para las bases de datos del estudio más sencillas; para las bases de datos complejas, hace falta crear una base de datos relacional utilizando programas de gestión de bases de datos basados en el lenguaje de consulta estructurado (*Structured Query Language*, SQL).
- Las consultas a la base de datos clasifican y filtran los datos, y calculan valores basados en los campos de datos. Se usan las consultas para supervisar la introducción de los datos, realizar informes sobre el progreso del estudio y dar formato a los resultados para su análisis.
- La pérdida de la base de datos debe evitarse mediante copias de seguridad frecuentes y depósito externo, y mediante el archivo

de copias de versiones clave de la base de datos para su uso futuro.

Volver al principio

BIBLIOGRAFÍA

1. Prokscha S. Practical guide to clinical data management, 3rd ed. Boca Raton: CRC Press, 2012. Citado aquí

2. Newman TB, Liljestrand P, Jeremy RJ, et al. Outcomes among newborns with total serum bilirubin levels of 25 mg per deciliter or more. N Engl J Med 2006;354(18):1889-1900. Citado aquí

3. Codd EF. A relational model of data for large shared data banks. Communications of the ACM 1970;13(6):377-387. Citado aquí

4. Date CJ. An introduction to database systems, 7th ed. Reading, Mass: Addison-Wesley, 2000. Citado aquí

5. Grinnon ST, Miller K, Marler JR, et al. National Institute of Neurological Disorders and Stroke common data element project—approach and methods. Clin Trials 2012;9(3):322-329. Citado aquí

6. NCI. The National Cancer Institute Cancer Data Standards Registry and Repository. 2012. Available from: https://cabig.nci.nih.gov/concepts/caDSR/, accessed 9/29/12. Citado aquí

7. FDA. Driving biomedical innovation: initiatives to improve products for patients. October, 2011. Available from: http://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UCM274464.pdf, accessed 1/29/13. Citado aquí

8. CDISC. The Clinical Data Interchange Standards Consortium Study data tabulation model. 2012. Available from: http://www.cdisc.org/sdtm, accessed 1/29/2013. Citado aquí

9. DHHS. Guidance for industry: computerized systems used in clinical trials. May, 2007. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070266.pdf, accessed 1/29/2013. Citado aquí

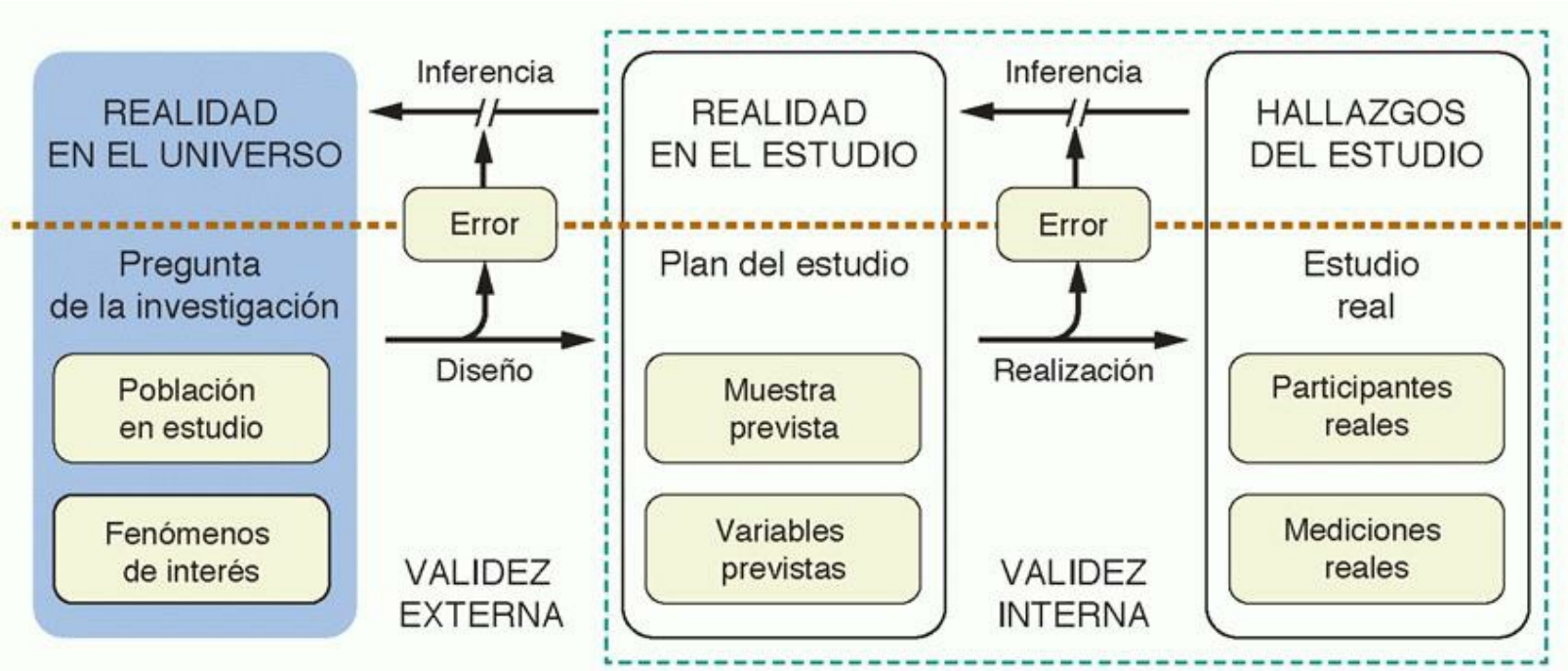
10. DHHS. Protecting personal health information in research: understanding the HIPAA Privacy Rule. 2003. Available from: http://privacyruleandresearch.nih.gov/pr_02.asp, accessed 1/29/2013. Citado aquí

CAPÍTULO 17 Realización del estudio y control de calidad
Deborah G. Grady
Stephen B. Hulley

La mayor parte de este libro ha tratado sobre el lado izquierdo del modelo de la investigación clínica, en el que se abordan asuntos del diseño (fig. 17-1). En este capítulo nos dirigimos hacia al segundo miembro, la realización. Incluso el mejor de los planes concienzudamente elaborado en el sillón puede resultar diferente en la práctica. Quizá no se disponga de personal de investigación experimentado, el espacio para realizar el estudio puede ser inferior al óptimo, los participantes pueden estar menos deseosos de participar de lo que se esperaba, la intervención puede tolerarse mal y las mediciones resultar difíciles. Las conclusiones de un estudio bien diseñado pueden echarse a perder por ignorancia, descuido, falta de formación y de normalización, y otros errores en la finalización y la realización del protocolo.

El éxito en la realización de un estudio empieza con la reunión de los recursos, entre ellos el espacio, el personal y la gestión económica para poner en marcha el estudio. La siguiente tarea es finalizar el protocolo mediante un proceso de evaluación preliminar de los planes de inclusión, medición e intervención en un intento de evitar la necesidad de revisiones del protocolo después de que haya comenzado la recogida de los datos. Después, el estudio se ejecuta con un método sistemático de control de calidad de los procedimientos clínicos y de laboratorio y del manejo de los datos, siguiendo los principios de las prácticas clínicas correctas (PCC) respaldadas por la FDA.

Algunas de las estrategias de este capítulo se refieren a estudios extensos con grandes equipos de investigación distribuidos en múltiples centros, que son dirigidos por investigadores con experiencia. Sin embargo, esta información también se aplica a investigadores principales que pueden ser coinvestigadores de estos estudios extensos, o investigadores principales en un estudio de menor tamaño.



▪FIGURA 17-1. Este capítulo se centra en el área que está dentro de la línea verde discontinua: realización de un proyecto de investigación.
Herramientas de imágenes

▪ REUNIÓN DE LOS RECURSOS

Espacio

Es posible realizar algunos estudios clínicos totalmente en Internet, utilizando sistemas interactivos basados en la web, intervenciones enviadas por correo (como fármacos o dispositivos), supervisión a distancia, visitas domiciliarias para las mediciones e introducción de datos en Internet. Sin embargo, la mayoría de los estudios de investigación siguen necesitando espacio físico para realizar las visitas del estudio y hacer las mediciones. Este espacio debe ser accesible, atractivo y suficiente. Si no se consigue negociar con éxito un espacio al principio del proceso de planificación del estudio, pueden surgir dificultades para conseguir participantes, puede haber un escaso cumplimiento de las visitas del estudio, los datos pueden quedar incompletos y el personal no trabajará bien. El espacio para la investigación clínica debe tener un acceso fácil para los participantes y contar con un aparcamiento adecuado disponible. Debe ser acogedor, cómodo y lo suficientemente amplio como para que quepa el personal, el equipo para las determinaciones y los fármacos para el estudio y los archivos relacionados con este. Si hay que realizar una exploración física, debe disponerse de una zona privada y un lugar para lavarse las manos. Si los participantes deben ir a otros lugares para hacerse pruebas (como el laboratorio del hospital o el servicio

de radiología), también deben tener un fácil acceso. En algunos estudios, como los que se realizan en pacientes enfermos o que suponen la realización de intervenciones que pudieran ser peligrosas, puede ser necesario contar con acceso a equipos de reanimación cardiopulmonar.

Muchos centros médicos académicos cuentan con centros de investigación clínica que proporcionan un espacio totalmente equipado para la investigación y personal de investigación experimentado. Los centros de investigación clínica a menudo permiten realizar determinaciones especializadas (como ingesta calórica, densidad ósea y estudios de «pinzamiento» de insulina), y pueden proporcionar acceso a otros servicios (como inclusión de los participantes, gestión de bases de datos y análisis estadístico). Son centros que ofrecen una excelente opción para realizar investigaciones clínicas y aplicadas, pero, generalmente, precisan procedimientos separados de solicitud y revisión, y reembolso por los servicios.

Equipo de la investigación

Los equipos de investigación oscilan, en cuanto a tamaño, desde pequeños (a menudo solo el investigador y un ayudante a tiempo parcial) hasta contar con un personal abundante a jornada completa para los grandes estudios. Independientemente del tamaño, todos los equipos de investigación deben realizar actividades similares y cumplir las mismas funciones, que se describen en la tabla 17-1 . Con frecuencia una persona desempeña varias de estas actividades. Sin embargo, algunos de estos deberes precisan una experiencia especial, como el análisis y la programación estadísticos. Algunos miembros del equipo, como los directores económicos y de recursos humanos, suelen trabajar para la universidad o el centro médico, y habrán sido proporcionados por la unidad o el departamento del investigador. Independientemente del tamaño del equipo del estudio, el investigador principal (IP) debe asegurarse de que se realicen todas las funciones que se describen en la tabla 17-1 .

Tras decidir el número de miembros que van a componer el equipo y distribuir las tareas, el siguiente paso será trabajar con un administrador del departamento para encontrar candidatos para el trabajo cualificados y con experiencia. Puede resultar difícil, porque la formación para algunos de los miembros del equipo es variable y los requisitos laborales variarán de un estudio a otro. Por ejemplo, la posición esencial de director del proyecto la puede ocupar una persona con experiencia en enfermería, farmacia, salud pública, servicios de laboratorio o investigación farmacéutica, y las obligaciones de este puesto pueden variar mucho.

La mayor parte de las universidades y centros médicos cuentan con métodos formales para anunciar ofertas laborales, pero también pueden ser útiles otras vías, como anuncios en periódicos y en la web. El método más seguro es encontrar personal cuya competencia se conozca, como, por ejemplo, alguien que trabaje con un compañero cuyo proyecto haya finalizado. También es habitual negociar con los compañeros para contratar a su personal con experiencia a tiempo parcial. Algunos centros médicos académicos o unidades del centro médico tienen un conjunto de coordinadores de investigación con experiencia y otro personal al que se puede contratar a tiempo parcial.

PAPEL	FUNCIÓN	COMENTARIOS
Investigador principal	Responsable final del diseño, financiación, dotación de personal, dirección y calidad del estudio, y de comunicar los hallazgos	
Director de proyecto/coordinador clínico	Proporciona la gestión día a día de todas las actividades del estudio	Con experiencia, responsable, meticuloso y con buenas aptitudes interpersonales y organizativas
Encargado de la inclusión	Asegura que participe el número deseado de pacientes elegibles	Informado y con experiencia en diversas técnicas de inclusión
Auxiliar de investigación/personal clínico	Realiza los procedimientos y las mediciones	La exploración física y otros procedimientos especializados pueden precisar titulaciones especiales
Coordinador del	Asegura que todo el personal sigue los	Observa los procedimientos del estudio y asegura el

control de calidad	procedimientos habituales y supervisa el control de calidad	cumplimiento de los procedimientos operativos estándar, puede supervisar auditorías por grupos externos, como la Food and Drug Administration (FDA) de Estados Unidos
Gestor de los datos	Diseña, evalúa y aplica el sistema de introducción, edición y gestión de los datos	
Programador/analista	Elabora informes del estudio que describen la inclusión, el cumplimiento y la calidad de los datos; realiza análisis de los datos	Trabaja bajo la supervisión del investigador principal (IP) y el estadístico
Estadístico	Colabora en el diseño del estudio, calcula la potencia y el tamaño de la muestra, diseña el plan de análisis y las directrices para la supervisión de los datos y la seguridad, e interpreta los hallazgos	Con frecuencia tiene un papel importante en el diseño general, la realización, el seguimiento intermedio, el análisis de los datos y la presentación de los resultados del estudio
Auxiliar administrativo	Proporciona apoyo administrativo, convoca reuniones, etc.	
Director económico	Prepara el presupuesto y gestiona los gastos	Realiza proyecciones para ayudar a gestionar el presupuesto
Director de recursos humanos	Ayuda a preparar las descripciones del trabajo, las contrataciones y las evaluaciones	Ayuda a gestionar los asuntos y problemas del personal

*En estudios pequeños, una persona puede adoptar varios de estos papeles; otros, como los gestores de recursos económicos y humanos, generalmente proceden del departamento y se comparten con otros miembros docentes.

Dirección y formación del equipo

La calidad de un estudio en el que participa más de una persona en el equipo investigador empieza con la integridad y el liderazgo del IP. Este debe garantizar que todo el personal cuente con la formación y la certificación adecuadas para realizar sus funciones. Debe transmitir claramente el mensaje de que son primordiales la protección de los participantes humanos, el mantenimiento de la privacidad, la integridad y exactitud de los datos, y la presentación adecuada de los hallazgos de la investigación. No puede supervisar todas las mediciones realizadas por los compañeros y por el personal, pero, si da la sensación de que, en general, confía en todas las actividades del estudio y se preocupa mucho por la protección de los participantes humanos y la calidad de los datos, la mayoría de las personas corresponderán de igual forma. Es útil reunirse con todos los miembros del equipo de vez en cuando para expresar el reconocimiento y comentar problemas y soluciones. Un buen jefe sabe delegar autoridad de forma adecuada y, al mismo tiempo, establecer un sistema jerárquico de supervisión que asegure un control suficiente de todos los aspectos del estudio.

Desde el principio de la fase de planificación, el investigador debe coordinar reuniones del personal periódicamente, con todos los miembros del equipo. La agenda de las reuniones debe distribuirse por adelantado, con los informes de progreso por las personas a las que se ha otorgado responsabilidad en áreas específicas del estudio. Estas reuniones son una oportunidad para descubrir y solucionar problemas, y para que todo el mundo intervenga en el proceso de desarrollo del proyecto y en la realización de la investigación. Las reuniones mejoran con exposiciones científicas y actualizaciones relacionadas con el proyecto. Son una buena fuente de moral e interés en los objetivos del estudio, y proporcionan formación sin abandonar el trabajo.

La mayor parte de las universidades y centros médicos orientados a la investigación proporcionan gran número de recursos institucionales para realizar la investigación clínica, como recursos humanos y servicios de gestión económica, servicios de consultoría, y centros de investigación clínica centralizados que proporcionan espacio y personal con experiencia. Muchas universidades cuentan también con laboratorios centrales en los que pueden realizarse determinaciones especializadas, un espacio y un equipo centralizados

para almacenar imágenes o muestras biológicas, servicios centralizados de gestión de bases de datos, centros de inclusión profesionales, experiencia con la U.S. Food and Drug Administration (FDA) y las normas de otras agencias de registro sanitario, y bibliotecas de formularios y documentos de estudios. Esta infraestructura puede no estar claramente a la vista en un gran centro, y los investigadores deberán procurar familiarizarse con sus recursos locales antes de intentar hacerlo por sí mismos.

Puesta en marcha del estudio

Al principio del estudio, el IP debe finalizar el presupuesto, elaborar y firmar los contratos que se precisen, definir los cargos del personal, contratar y formar al personal, obtener la aprobación del CEIC, redactar el manual de instrucciones, desarrollar y verificar formularios y cuestionarios, desarrollar y evaluar la base de datos, y planificar la inclusión de los participantes. A este período de actividad del estudio antes de la inclusión del primer participante se le denomina puesta en marcha del estudio, y precisa un intenso esfuerzo. El tiempo y la planificación adecuados para poner en marcha el estudio son importantes para poder realizar un estudio de gran calidad.

Algo esencial será contar con una financiación adecuada para poder realizar el estudio. El presupuesto se habrá preparado al enviar la propuesta de financiación, antes de iniciar el estudio (cap. 19). La mayor parte de las universidades y centros médicos emplean personal con experiencia para ayudar al desarrollo de los presupuestos (gerente previo a la adjudicación). Es una buena idea conocer bien a esta persona, respetar su nivel de estrés en relación con las fechas límite cumpliendo los objetivos del calendario, y comprender las regulaciones relacionadas con las diversas fuentes de financiación.

En general, las normas para gastar los fondos del NIH y otros fondos públicos son considerablemente más restrictivas que las de los fondos de la industria o de las fundaciones. La cantidad total del presupuesto no puede aumentarse si el trabajo se encarece más de lo previsto, y el trasvase de dinero entre partidas de gastos (p. ej., personal, equipo, suministros, viajes) o las reducciones importantes del esfuerzo porcentual del personal principal, en general, necesitan la aprobación del promotor. Las universidades y centros médicos suelen contratar a personal especializado cuya principal responsabilidad es asegurar que los fondos disponibles para un investigador a través de subvenciones y contratos se gastan adecuadamente. Este gerente posterior a la adjudicación debe preparar informes y proyecciones regulares que permitan que el investigador realice ajustes en el presupuesto, para poder hacer así el mejor uso de los fondos disponibles durante el tiempo que dure el estudio, garantizando que no se haya gastado todo el presupuesto al final del estudio. Tener un superávit pequeño al final del estudio puede ser algo bueno, porque los promotores muchas veces autorizan «extensiones sin coste» que permiten el uso de los fondos sobrantes después de la finalización formal del período del estudio para finalizar o ampliar el trabajo descrito en el ámbito de la subvención.

El presupuesto de un estudio patrocinado por una compañía farmacéutica forma parte de un contrato que incorpora el protocolo y un esquema claro de las tareas que van a desarrollar el investigador y el promotor. Los contratos son documentos legales que obligan al investigador a realizar las actividades, y describen el plazo y el importe de pago para «productos» especificados, como cumplir los hitos de la inclusión y enviar informes de progreso. En las universidades o centros médicos se necesitan abogados para ayudar a elaborar esos contratos, y para asegurar que se protegen los derechos de propiedad intelectual del investigador, el acceso a los datos, los derechos de publicación, etc. Sin embargo, los abogados generalmente no están familiarizados con las tareas necesarias para realizar un estudio específico, y la participación del investigador es esencial, especialmente en lo que respecta al ámbito del trabajo y los productos.

Aprobación por el comité ético de investigación clínica

El CEIC debe aprobar el protocolo del estudio, el formulario de consentimiento y los materiales para reunir a los participantes antes de que comience la inclusión (cap. 14). Los investigadores deben estar familiarizados con los requisitos de su CEIC local y del tiempo necesario para lograr la aprobación. El personal del CEIC suele ser muy útil en estos temas, y debe contactarse con ellos pronto para comentar cualquier aspecto del procedimiento y de las decisiones del diseño que afecte a los participantes del estudio.

Desarrollo del manual de instrucciones y de formularios

El protocolo del estudio suele ampliarse para crear el manual de instrucciones, que contiene el protocolo, la información sobre la organización y las políticas del estudio, y una versión detallada de la sección de metodología del protocolo del estudio (apéndice 17A). Especifica exactamente cómo reunir e incluir a los participantes en el estudio, y describe todas las actividades que se realizan en cada consulta: cómo se conseguirá la aleatorización y el enmascaramiento, cómo se medirá cada variable, los procedimientos del control de calidad, las prácticas sobre la gestión de los datos, el plan de análisis estadístico, y el plan para la supervisión de los datos y la seguridad (cap. 11). También debe incluir todos los cuestionarios y formularios que se utilizarán en el estudio, con instrucciones sobre el contacto con los participantes del estudio, la realización de entrevistas, cómo rellenar y codificar los formularios del estudio, la introducción y la edición de datos, y la recopilación y el procesamiento de las muestras. Un manual de instrucciones es algo esencial cuando la investigación la realizan varias personas, fundamentalmente cuando hay una colaboración entre investigadores que se encuentran en más de un lugar. Incluso aunque un solo investigador realice todo el trabajo por sí mismo, las definiciones por escrito de las operaciones ayudarán a reducir la variación y los cambios aleatorios en la técnica de medición que pueden producirse con el tiempo.

El diseño de formularios para la recogida de los datos tendrá una importante influencia en la calidad de los datos y en el éxito del estudio (cap. 16). Antes de que se consiga el primer participante, los formularios tienen que haberse verificado. Toda entrada en un formulario que conlleve un juicio precisará definiciones explícitas, que deben resumirse brevemente en el propio formulario y explicarse con más detalle en el manual de instrucciones. Las preguntas deben ser coherentes, y el formato de su secuencia debe estar claro, con patrones de exclusión (v. apéndice 15). La verificación previa garantizará la claridad y la facilidad de uso. Marcar cada página con la fecha, el nombre y el número de ID del participante y del personal garantiza la integridad de los datos. Antes del inicio del estudio deben verificarse los formularios digitales basados en Internet, los ordenadores portátiles, asistentes personales digitales y otros dispositivos para obtener datos, y el manual de instrucciones contendrá las normas para su uso.

Diseño de la base de datos

Antes de incluir al primer participante, debe crearse, almacenarse, comprobarse y analizarse la base de datos que se usará para introducir, almacenar, actualizar y supervisar los datos. Dependiendo del tipo de bases de datos que se vaya a utilizar y del ámbito del estudio, el desarrollo y la verificación del sistema de introducción y gestión de los datos pueden necesitar semanas o meses después de que se haya identificado, contratado y formado al personal con las aptitudes adecuadas para ello. Muchos centros médicos académicos ofrecen servicios para ayudar a los investigadores a elaborar una base de datos adecuada, y programas informáticos de bases de datos de utilización general. Para estudios muy extensos se dispone de servicios de diseño y gestión de bases de datos profesionales, aunque es bueno pedir consejo sobre estas opciones a expertos técnicos del centro que sean de confianza y a asesores con mayor experiencia.

Incluso para estudios pequeños, el tiempo que se invierte al comienzo de la creación de una base de datos que albergará los datos del estudio habitualmente es un tiempo bien invertido (cap. 16). Los investigadores ansiosos por comenzar un estudio y empezar a registrar datos, en ocasiones registran los datos en formularios de papel o en una hoja de cálculo como Microsoft Excel, en lugar del programa real de la base de datos. Aunque este abordaje es más fácil al comienzo, acaba costando mucho más tiempo y esfuerzo posteriormente, cuando llega el momento de analizar los datos. La ventaja de configurar una base de datos en las primeras fases es que permite que el investigador considere al inicio qué valores de cada una de las variables son aceptables, y que prohíba o genere alertas para los valores que estén fuera del intervalo, que sean ilógicos o que estén ausentes. Los sistemas de introducción y gestión de los datos de calidad elevada mejoran el control de calidad en el momento de la recogida de los datos o su introducción, y reducen el tiempo que será necesario posteriormente para realizar la limpieza de los datos. No obstante, la principal ventaja de un sistema de datos de calidad elevada es evitar descubrir en fases tardías del estudio que hay un gran número de valores ausentes, fuera del intervalo o ilógicos que no se pueden corregir.

Inclusión

En el capítulo 3 se describen los métodos para conseguir el número propuesto de participantes para un estudio. Deseamos destacar aquí que conseguir los participantes a tiempo es el aspecto más difícil de muchos estudios. El momento adecuado, el personal, los recursos y la experiencia son esenciales para poder lograrlo, y debe planificarse todo bien antes del inicio del estudio.

Volver al principio

▪ FINALIZACIÓN DEL PROTOCOLO

Pruebas preliminares y «ensayos generales»

Las pruebas preliminares y los estudios piloto se diseñan para evaluar la viabilidad, la eficacia y el coste de los métodos del estudio, la reproducibilidad y la exactitud de las mediciones; las probables tasas de reclutamiento; y (en ocasiones) las tasas de respuesta y las magnitudes del efecto. La naturaleza y la escala de las pruebas preliminares y los estudios piloto dependen del diseño del estudio y de sus necesidades. En la mayor parte de los estudios, una serie de pruebas preliminares o un pequeño estudio piloto serán adecuados, pero, para los estudios caros y de gran tamaño, lo adecuado puede ser un estudio piloto a gran escala. Quizá sea deseable gastar hasta el 1 % del coste final del estudio en asegurarse de que las estrategias para conseguir participantes funcionarán, de que las mediciones son las adecuadas y de que los cálculos del tamaño de la muestra son realistas.

Las pruebas preliminares son evaluaciones de cuestionarios, mediciones o procedimientos específicos que puede realizar el personal del estudio para valorar su funcionalidad, idoneidad y viabilidad. Por ejemplo, la prueba preliminar de la introducción de datos y del sistema de gestión de la base de datos suele realizarse proporcionando al personal del estudio formularios completos con datos perdidos, fuera de los límites o ilógicos, introduciendo a continuación estos datos, y realizando después una prueba para asegurarse de que el sistema de edición de datos identifica estos errores.

Antes de iniciar el estudio, será una buena idea poner a prueba los planes de las visitas en la consulta y otros procedimientos del estudio

en un ensayo general. El objetivo es subsanar problemas con la última serie de instrumentos y procedimientos. Lo que sobre el papel parece ser un protocolo sin problemas suele mostrar problemas logísticos e importantes en la práctica, y el ensayo general conseguirá mejoras en el método. El propio investigador principal puede actuar como participante simulado para experimentar el estudio y el equipo de investigación desde ese punto de vista.

Revisiones menores del protocolo una vez iniciada la recopilación de datos

Independientemente del cuidado y la atención con que se haya diseñado el estudio y se hayan probado los procedimientos, los problemas aparecerán inevitablemente una vez que el estudio haya comenzado. La norma general es realizar el menor número de cambios posibles en ese momento. A veces, no obstante, las modificaciones del protocolo pueden reforzar el estudio.

La decisión sobre si un cambio menor mejorará la integridad del estudio es, a menudo, una concesión mutua entre el beneficio que se logra con la mejora de la metodología y los inconvenientes que supone alterar la uniformidad de los métodos del estudio, el gasto de tiempo y dinero para cambiar el sistema, y la introducción de confusión en algunos de los miembros del equipo. Las decisiones que tan solo conllevan hacer más específica una definición operativa son relativamente sencillas. Por ejemplo, en un estudio que excluye a personas con abuso de alcohol, ¿puede incluirse a una persona que ha mantenido la abstinencia durante varios años? La decisión debe tomarse consultando con los coinvestigadores, pero con la comunicación adecuada a través de memorandos y el manual de instrucciones para asegurarse de que todo el personal lo aplique uniformemente durante el resto del estudio. Con frecuencia, los ajustes menores de este tipo no necesitan la aprobación del CEIC, sobre todo si no suponen cambiar el protocolo que ya ha sido aprobado, pero el IP debe consultar a un miembro del CEIC por si hubiera alguna duda. Deben identificarse todos los cambios del protocolo, de los formularios de consentimiento informado, del manual de instrucciones o de otros documentos del estudio asignando al documento revisado un número de versión nuevo, y se deben aplicar métodos para garantizar que se utilice la última versión de cada uno de los documentos.

Revisiones importantes del protocolo una vez iniciada la recogida de datos

Los cambios importantes en el protocolo del estudio, como incluir diferentes tipos de participantes o cambiar la intervención o la variable de respuesta, constituyen un problema grave. Aunque puede haber buenas razones para realizar estos cambios, deben hacerse con la intención de analizar y comunicar los datos por separado si esto conduce a una interpretación más adecuada de los hallazgos. Los juicios se ilustran con dos ejemplos del ensayo *Raloxifene Use for the Heart* (RUTH), un ensayo clínico multicéntrico del efecto del tratamiento con raloxifeno sobre los episodios coronarios en 10 101 mujeres con riesgo elevado de sufrir episodios coronarios. La definición inicial de la respuesta primaria fue la aparición de infarto de miocardio (IM) no mortal o muerte por cardiopatía isquémica. Al principio del ensayo se observó que la incidencia de esta respuesta era inferior a la esperada, probablemente debido a que nuevas cointervenciones clínicas, como la trombólisis y la angioplastia coronaria, reducían el riesgo de IM. Tras una cuidadosa consideración, el comité ejecutivo del estudio RUTH decidió cambiar la variable de respuesta principal para incluir síndromes coronarios agudos distintos al IM. Este cambio se realizó al principio del ensayo; se había recogido información adecuada sobre los posibles episodios cardíacos para determinar si cumplían los criterios de síndrome coronario agudo, permitiendo que se buscaran en la base de datos del estudio episodios coronarios agudos que se habían producido antes de realizarse el cambio (1).

También al principio del estudio RUTH, los resultados que surgieron del ensayo *Multiple Outcomes of Raloxifene Evaluation* (MORE) demostraron que el riesgo relativo de sufrir cáncer de mama se reducía notablemente con el tratamiento con raloxifeno (2). Estos resultados no eran concluyentes, ya que el número de cánceres de mama era pequeño, y había dudas sobre la posibilidad de generalizar, ya que todas las mujeres que participaron en el estudio MORE tenían osteoporosis. Para determinar si el raloxifeno reduciría también el riesgo de cáncer de mama en otra población (mujeres mayores sin osteoporosis), el comité ejecutivo del estudio RUTH decidió añadir el cáncer de mama como segunda respuesta principal (1).

Cada uno de estos cambios era importante, y fue necesaria una enmienda al protocolo, la aprobación del CEIC de cada uno de los centros clínicos y la aprobación de la FDA, y la revisión de un gran número de formularios y documentos del estudio. Estos son ejemplos de revisiones importantes que aumentan la viabilidad o el contenido de la información del estudio sin comprometer su integridad global. Cambiar el protocolo no siempre tiene tanto éxito. Las revisiones importantes solo deben realizarse tras sopesar los pros y los contras con miembros del equipo de investigación y asesores adecuados, como el comité de supervisión de los datos y la seguridad, el promotor y la agencia que financia el estudio. El investigador debe luego afrontar las posibles consecuencias del cambio cuando analice los datos y extraiga las conclusiones del estudio.

Cierre

En algún punto de todos los estudios longitudinales y ensayos clínicos se detiene el seguimiento de los participantes. El período durante el cual los participantes realizan la última consulta del estudio se denomina «cierre» y presenta varios puntos que se deben planificar (3). Como mínimo, en la última consulta, el personal debe agradecer a los participantes su tiempo y su esfuerzo, e informarles de que su participación fue esencial para que el estudio tuviera éxito. Además, este período de cierre puede incluir las actividades siguientes:

- Generalmente, se debe informar a los participantes (y a sus médicos) de los resultados de las pruebas de laboratorio con importancia clínica y de otras determinaciones que se realizaron durante el estudio, ya sea en persona, en la última consulta (con una copia por escrito) o, posteriormente, por correo.

- En un ensayo clínico con enmascaramiento debe comunicarse a los participantes el tratamiento que han recibido, ya sea en la última consulta o por correo, cuando todos los participantes hayan finalizado el ensayo y se hayan realizado los principales análisis de los datos o se haya publicado el principal manuscrito basado en los resultados del estudio.

- En general, se debe enviar por correo a los participantes (y a sus médicos) una copia del manuscrito principal basado en los resultados del estudio u otra descripción de los hallazgos, redactado en un lenguaje no técnico, cuando se realice la presentación o la publicación, con un número de teléfono para los participantes que tengan preguntas.

- Una vez que todos los participantes hayan finalizado el estudio, se les puede invitar a una recepción durante la que el IP les dará las gracias, comentará los resultados del estudio y responderá a sus preguntas.

Volver al principio

▪ CONTROL DE CALIDAD DURANTE EL ESTUDIO

Prácticas clínicas correctas

Un aspecto esencial de la investigación clínica es el método para asegurar que todos los aspectos del estudio sean de la mayor calidad. Las directrices para una investigación de calidad, denominadas prácticas clínicas correctas (PCC), se desarrollaron para aplicarse específicamente a los ensayos clínicos en los que se estudian fármacos y que precisan la aprobación de la FDA u otros organismos reguladores, y se definen como «un patrón de calidad ético y científico internacional para el diseño, la realización, el registro y la notificación de estudios que suponen la participación de seres humanos. El cumplimiento de esta norma ofrece una garantía pública de que se protegen los derechos, la seguridad y el bienestar de los participantes en el estudio» (4).

Estos principios se han aplicado cada vez más en ensayos clínicos patrocinados por organismos federales y otros organismos públicos, y en diseños de investigación distintos de los ensayos (tabla 17-2). Los requisitos de las PCC se describen con detalle en el título 21 del código de regulaciones federales de la FDA (4 , 5). La Conferencia Internacional de Armonización (6) proporciona directrices para el control de calidad que usan los organismos reguladores en Europa, Estados Unidos y Japón.

La práctica clínica correcta se consigue mejor mediante procedimientos operativos estándar (POE) para todas las actividades relacionadas con el estudio. Puede considerarse que el protocolo, el plan del análisis estadístico y el plan de supervisión de los datos y la seguridad son POE, pero, a menudo, no abarcan áreas del tipo de cómo se ha formado el personal, cómo se ha desarrollado y verificado la base de datos, o cómo se han mantenido los archivos del estudio, se ha preservado su confidencialidad y se han copiado los archivos del estudio. Muchos centros médicos académicos tienen personal que se especializa en procesos para cumplir con las directrices de PCC, y pueden suministrar diversas plantillas y modelos de POE. En el capítulo 14 se comenta el tema relacionado de la conducta ética de la investigación, y en este capítulo nos centramos en el control de calidad de los procedimientos del estudio y de la gestión de los datos.

- El diseño está respaldado por datos preclínicos, en animales y otros datos cuando proceda
- El estudio se realiza según los principios éticos de la investigación
- Se sigue cuidadosamente un protocolo por escrito
- Los investigadores y los profesionales que prestan asistencia médica están formados y cualificados
- Todos los procedimientos clínicos y de laboratorio cumplen patrones de calidad
- Los datos son fiables y exactos
- Se mantienen unos registros completos y exactos
- Los métodos estadísticos se especifican previamente y se siguen meticulosamente

- Los resultados se comunican clara y objetivamente

Control de calidad de los procedimientos clínicos

Es una buena idea que uno de los miembros del equipo investigador sea el coordinador del control de calidad, que es el responsable de realizar las técnicas adecuadas de control de calidad en todos los aspectos del estudio, supervisar la formación del personal y realizar un seguimiento del uso de los procedimientos de control de calidad durante el estudio. El objetivo es detectar posibles problemas antes de que se produzcan, y evitarlos. El coordinador del control de calidad puede ser también responsable de preparar lo necesario para las auditorías del CEIC, la FDA, el patrocinador del estudio o el NIH, y de actuar como persona de contacto para dichas auditorías. El control de calidad empieza durante la fase de planificación y continúa durante todo el estudio (tabla 17-3).

- Manual de instrucciones. El manual de instrucciones es un aspecto muy importante del control de calidad (v. apéndice 17A). Para ilustrarlo, considere la altura en un estudio en el que el cambio de altura se utilizará como factor predictivo de osteoporosis. Como la medición de la altura es una variable de respuesta parcialmente subjetiva, para la cual no hay un método de referencia viable, el manual de instrucciones debe mostrar instrucciones específicas sobre el tipo de dispositivo de medición a utilizar (marca y modelo del estadio metro), además de las instrucciones para preparar al participante para la medición (quitar el calzado), la colocación del paciente en el dispositivo de medición y la toma de la medición.

- Calibración, formación y certificación. Se debe realizar una calibración profesional de los dispositivos de medición (básculas, estadiómetros, equipos de imagen, equipos de laboratorio, etc.) antes del comienzo del estudio. La formación estandarizada del personal del estudio es esencial para una investigación de calidad elevada. Todo el personal implicado en el estudio debe recibir una formación adecuada antes del comienzo del estudio, y se debe certificar su competencia en relación con procedimientos y mediciones fundamentales. Por ejemplo, en relación con la medición de la altura, se puede formar a los miembros del equipo en todos los aspectos de la medición, y deberán obtener mediciones satisfactorias en participantes simulados cuya altura se conozca. El procedimiento de certificación se debe complementar durante el estudio por nuevas certificaciones programadas, y en el centro de investigación se debe mantener un registro de formación, certificación y renovación de la certificación.

- Revisión del rendimiento. Los supervisores deben revisar cómo se realizan los procedimientos clínicos, asistiendo a consultas clínicas representativas o mediante llamadas telefónicas. Tras obtener la autorización del participante en el estudio, el supervisor puede estar presente en al menos un ejemplo completo de cada tipo de entrevista y procedimiento técnico que realice cada miembro de este equipo de investigación. Al principio puede parecer embarazoso, pero pronto será cómodo. Es útil utilizar una lista de verificación normalizada (proporcionada por adelantado, y basada en el protocolo y el manual de instrucciones) durante estas observaciones. Después puede facilitarse la comunicación entre el supervisor y el miembro del equipo revisando la lista y resolviendo, de un modo positivo y no peyorativo, todos los puntos sobre el control de calidad que se anotaron. El momento y los resultados de las revisiones de la realización deben anotarse en los registros de formación.

Pasos que preceden al estudio	Desarrollar un manual de instrucciones
	Definir estrategias de inclusión
	Crear definiciones operativas de las mediciones
	Crear instrumentos y formularios normalizados
	Crear sistemas de control de calidad
	Crear sistemas de enmascaramiento de participantes e investigadores
	Designar un coordinador del control de calidad
	Formar al equipo investigador y documentarlo
	Certificar al equipo investigador y documentarlo
	Proporcionar una dirección uniforme y atenta
	Mantener reuniones frecuentes con el personal

Pasos durante el estudio	Crear procedimientos especiales para intervenciones con fármacos
	Certificar de nuevo al equipo investigador
	Realizar revisiones periódicas del rendimiento
	Comparar las mediciones entre los técnicos y a lo largo del tiempo

*Son procedimientos clínicos: la medición de la presión arterial, la entrevista estructurada, la revisión de historias, etc.

Es útil que participen compañeros del equipo de investigación como revisores para construir un trabajo en equipo, así como para asegurar la aplicación constante de métodos normalizados entre los miembros del equipo que realizan la misma tarea. Una ventaja de usar compañeros como observadores en este sistema es que todos los miembros del equipo investigador adquieren una sensación de propiedad del proceso de control de calidad. Otra de las ventajas es que, a menudo, el observador aprende tanto observando la realización de otra persona como esta al final del proceso de revisión.

- Informes periódicos. Es importante tabular los datos de la calidad técnica de los procedimientos y las mediciones clínicas a intervalos regulares, ya que puede dar pistas sobre la presencia de mediciones ausentes, inexactas o variables. Las diferencias entre los miembros de un equipo de medición de la presión arterial en los niveles medios observados en los últimos 2 meses, por ejemplo, pueden descubrir diferencias en sus técnicas de medición. Igualmente, un cambio gradual en unos meses de la desviación típica de series de lecturas puede indicar que existe un cambio en la técnica de la medición. Los informes periódicos también deben hacer ver el éxito de la inclusión, la puntualidad de la introducción de datos, la proporción de variables ausentes y fuera del intervalo, el tiempo transcurrido hasta que se abordan las consultas sobre los datos, y el éxito del seguimiento y del cumplimiento de la intervención.

- Procedimientos especiales para intervenciones con fármacos. Los ensayos clínicos en los que se usan fármacos, sobre todo los que utilizan enmascaramiento, precisan una atención especial al control de calidad del etiquetado, la distribución y el almacenamiento del fármaco, la dispensación de la medicación, y la recopilación y la eliminación de la medicación no utilizada. Se asegura la provisión del fármaco y la dosis correctos mediante una planificación meticulosa con el fabricante o la farmacia de la investigación en relación con la naturaleza del método de distribución del fármaco, supervisando su aplicación y comprobando, en ocasiones, la composición de los medicamentos del estudio con enmascaramiento, para asegurarse de que contienen los componentes correctos. Los estudios farmacológicos también necesitan procedimientos claros y registros para seguir la recepción de medicamentos del estudio, el almacenamiento, la distribución y su devolución por los participantes.

Control de calidad para procedimientos de laboratorio

La calidad de los procedimientos de laboratorio puede controlarse usando muchos de los métodos descritos en la tabla 17-3 para los procedimientos clínicos. Además, el hecho de que las muestras se hayan extraído de los participantes (lo que crea la posibilidad de un etiquetado erróneo) y la naturaleza técnica de las pruebas analíticas conducen a varias estrategias especiales:

- Atención al etiquetado. Cuando la muestra de sangre de un participante se etiqueta de forma errónea con el nombre de otra persona, puede ser imposible corregir o incluso descubrir el error posteriormente. La única solución es la prevención, evitando los errores de etiquetado y los errores por transposición, comprobando cuidadosamente el nombre y el número del participante cuando se etiqueta cada muestra. Las copias generadas con ordenador de las etiquetas para los tubos de sangre y las historias aceleran el proceso de etiquetado y evitan los errores que pueden producirse cuando los números están escritos a mano. Un buen procedimiento al transferir suero de un tubo a otro consiste en etiquetar el otro tubo por adelantado, y sostener los dos tubos uno junto al otro, leyendo uno en voz alta mientras se comprueba el otro; esto puede hacerse también de forma automática con códigos de barras que se pueden escanear.

- Enmascaramiento. La tarea de mantener el enmascaramiento del observador es fácil cuando se trata de mediciones sobre muestras, y es siempre una buena idea marcar estas de modo que el técnico desconozca el grupo del estudio o el valor de otras variables esenciales. Incluso para procedimientos aparentemente objetivos, como una determinación automática de la glucemia, esta precaución reduce las oportunidades de que se produzca sesgo, y proporciona una sección de metodología más sólida al comunicar los resultados. Sin embargo, el enmascaramiento de la asignación al personal del laboratorio significa que tiene que haber procedimientos claros para comunicar los resultados anormales a un miembro del equipo que esté cualificado para revisar los resultados y decidir si debe notificarse al participante o debe emprenderse alguna otra acción. En los ensayos clínicos debe

haber también estrategias preparadas para desenmascarar (a veces urgentemente) si las determinaciones analíticas indican alteraciones que podrían asociarse a la intervención del ensayo y que requieren una acción inmediata.

- Duplicados con enmascaramiento, mezclas estándar y mediciones de consenso. Cuando se envían muestras o imágenes a un laboratorio central para su análisis químico o su interpretación, puede ser buena idea enviar duplicados con enmascaramiento (una segunda muestra de un subgrupo aleatorio de participantes a la que se asigna un número de ID separado y ficticio) mediante el mismo sistema. Esa estrategia da una medida de la precisión de la técnica de laboratorio. Otro método para muestras séricas que pueden guardarse congeladas es preparar una mezcla de suero al principio y enviar periódicamente alícuotas, a través del sistema, que están marcadas con números de ID ficticios y ocultando la asignación. Las medidas realizadas en la mezcla de suero al principio, mediante la mejor técnica disponible, establecen sus valores; se usa después la mezcla como un patrón de referencia durante el estudio, para obtener cálculos de exactitud y precisión. Un tercer método, para medidas que tienen variabilidad inherente, como la citología cervicovaginal y las lecturas de mamografías, es hacer participar a dos lectores independientes que desconozcan la asignación. Si ambos están de acuerdo dentro de los límites predefinidos, se establece el resultado. Los resultados discordantes pueden resolverse mediante discusión y consenso, o con la opinión de un tercer lector.

- Contratos con laboratorios comerciales. En algunos estudios, las mediciones biológicas realizadas con suero, sangre, células o tejidos se hacen bajo contrato con laboratorios comerciales. El laboratorio debe contar con la licencia y la certificación adecuadas, y en los archivos de las oficinas del estudio debe figurar una copia de estas certificaciones. Los laboratorios comerciales deben ofrecer datos sobre la reproducibilidad de sus mediciones, como los coeficientes de variación, garantizar un servicio puntual y disponer de procedimientos normalizados para manejar muestras codificadas, notificar los resultados anormales a los investigadores y transferir los datos a la base de datos principal.

Control de calidad de la gestión de los datos

El investigador debe establecer y comprobar previamente el sistema de gestión de los datos antes de que se inicie el estudio. Esto conlleva el diseño de formularios para registrar las determinaciones, la elección de los sistemas informáticos y los programas para la introducción, la edición y la gestión de los datos; el diseño de los parámetros de edición de los datos para entradas ausentes, fuera de los límites aceptables o ilógicos; la comprobación del sistema de gestión de los datos; y la planificación de tabulaciones ficticias para asegurar que se recogen las variables adecuadas (tabla 17-4).

Sea parco: recopile solo las variables necesarias

Seleccione el sistema y los programas informáticos adecuados para la gestión de la base de datos

Programe la base de datos para señalar los datos que se pierden, los que están fuera de los límites y los valores ilógicos

Estudie la base de datos con valores perdidos, fuera de límites e ilógicos

Planifique análisis y pruebas con tabulaciones simuladas

Diseñe formularios en papel o electrónicos que sean:

Autoexplicativos

Coherentes (p. ej., las opciones de múltiple elección son exhaustivas y mutuamente excluyentes)

Con un formato claro para la introducción de datos y flechas que dirigen patrones de exclusión

Impresos en minúsculas, usando las mayúsculas, el subrayado y la negrita para resaltar

Estéticos y de fácil lectura

Evaluados preliminarmente y validados (cap. 15)

Marcados en cada página con la fecha, el nombre, el número de ID o un código de barras

- Datos ausentes. Los datos ausentes pueden ser desastrosos si afectan a una gran proporción de las mediciones, e incluso la pérdida de unos pocos puede causar a veces un sesgo en las conclusiones. Un estudio de las secuelas a largo plazo de una operación que tiene una incidencia de mortalidad tardía del 5%, por ejemplo, podría infravalorar gravemente esta complicación si se perdiera el 1 % de los participantes durante el seguimiento y si el fallecimiento fuera una razón habitual para perderlos. Las conclusiones erróneas debidas a la pérdida de datos pueden corregirse a veces después del hecho (en este caso, mediante un intenso esfuerzo para seguir a los participantes perdidos), pero, a menudo, la medición no puede reemplazarse. Hay técnicas estadísticas para imputar valores perdidos según otras informaciones de la visita inicial o de otras visitas de seguimiento, o por los valores medios de los demás participantes. Aunque estas técnicas son útiles, particularmente en los análisis de múltiples variables en los que la acumulación de datos perdidos en muchas variables predictivas podría hacer que hubiera grandes proporciones de participantes no disponibles para el análisis, no garantizan unas conclusiones sin sesgo de ausencia de respuesta si existe una cantidad importante de observaciones perdidas.

  La única buena solución consiste en diseñar y realizar el estudio de un modo que evite la pérdida de datos, por ejemplo, disponiendo de un miembro del equipo investigador que compruebe los formularios de cumplimiento antes de que el participante salga de la consulta, diseñando interfaces de introducción de datos electrónicas que no permitan los saltos de entradas, y diseñando la base de datos de modo que se indique la pérdida de datos inmediatamente al personal del estudio (tabla 17-5). La pérdida de determinaciones clínicas debe abordarse mientras el participante se encuentra en la consulta, cuando es relativamente fácil corregir los errores que se hayan descubierto.

- Datos inexactos e imprecisos. Es este un problema insidioso que a menudo no se descubre, sobre todo cuando participa más de una persona en las mediciones. En el peor de los casos, el investigador diseña el estudio y deja la recogida de los datos a sus ayudantes. Cuando vuelve a analizar los datos, algunas de las mediciones pueden haber sufrido un importante sesgo por el uso constante de una técnica inadecuada. Este problema es particularmente grave cuando no pueden detectarse los errores en los datos después de que se hayan producido. El investigador supondrá que las variables indican lo que pretendía que indicaran e, ignorante del problema, puede extraer conclusiones erróneas de este estudio.

  La formación y la certificación del personal, la revisión periódica de los procedimientos y la evaluación frecuente de las diferencias en la media o el intervalo de los datos generados por diferentes miembros pueden ayudar a identificar o evitar estos problemas. La edición computarizada tiene un papel importante, al usar sistemas de introducción y gestión de datos programados para señalar o no permitir el envío de formularios con valores ausentes, incongruentes o fuera de los límites permitidos. Debe existir un procedimiento normalizado para cambiar los datos originales en cualquier formulario de datos. Generalmente, esto debe realizarse tan pronto como sea posible tras la recogida de los datos, y con un proceso consistente en marcar la entrada original (sin borrarla), firmar y fechar el cambio. Se deben incluir procesos similares en la introducción electrónica de los datos y los sistemas electrónicos de edición. Esto proporciona un «rastro de auditoría» para justificar los cambios en los datos y evitar el fraude.

Señalar o comprobar las omisiones y los errores importantes mientras los participantes permanecen en la consulta

Sin errores ni transposiciones del número de ID, el código del nombre o la fecha en todas las páginas

Se han rellenado todos los formularios correctos de la consulta especificada

No hay entradas ausentes ni patrones de exclusión defectuosos

Las entradas son legibles

Los valores de variables fundamentales están dentro de los límites permisibles

Los valores de variables fundamentales son coherentes entre sí (p. ej., edad y fecha de nacimiento)

Realizar mediciones periódicas de la varianza y de las distribuciones de frecuencia para descubrir valores anómalos

Generar periódicamente otras tabulaciones para descubrir errores (apéndice 17B)

La tabulación y la inspección periódicas de las distribuciones de frecuencia de las variables del estudio importantes a intervalos frecuentes permiten que el investigador valore la integridad y la calidad de los datos, en un momento en el que sigue siendo posible la corrección de errores anteriores (p. ej., estableciendo contacto con el participante por correo electrónico o por teléfono, o solicitando al participante que regrese al centro de investigación), y cuando pueden evitarse más errores en el resto del estudio. En el apéndice 17B se ofrece una lista útil de temas para los informes de control de calidad.

- Datos fraudulentos. Los investigadores clínicos que dirigen equipos de investigación deben tener en cuenta la posibilidad de que haya un compañero o empleado sin escrúpulos que decida inventar la información del estudio como mejor forma de tener el trabajo hecho. Los métodos para protegerse frente a un suceso de ese tipo consisten en tener un gran cuidado al elegir los compañeros y el personal, desarrollar una relación sólida con ellos para que todos entiendan explícitamente la conducta ética y la sigan estrictamente, estar alerta ante la posibilidad de fraude al examinar los datos, y hacer comprobaciones de improviso de la fuente primaria de los datos para asegurarse de que estos son reales.

Estudios cooperativos multicéntricos

Muchas preguntas de investigación precisan un mayor número de participantes de los que se dispone en un solo centro, y a menudo se abordan en estudios cooperativos llevados a cabo por equipos de investigación que trabajan en varios lugares. A veces se encuentran todos en la misma ciudad o estado, y un solo investigador puede supervisar a todos los equipos. Con frecuencia, sin embargo, este tipo de estudios los realizan investigadores que se encuentran en ciudades situadas a cientos de kilómetros, con diferentes fuentes de administración, estructuras administrativas y agencias de registro sanitario.

Los estudios multicéntricos de este tipo precisan pasos especiales para asegurar que todos los centros usen los mismos procedimientos y produzcan datos comparables que puedan combinarse en el análisis de los resultados. Un centro de coordinación establece una red de comunicación, coordina el desarrollo del manual de instrucciones, los formularios y otros aspectos normalizados del control de calidad del ensayo, forma al personal que realizará las mediciones en cada centro, y supervisa la gestión, el análisis y la publicación de los datos. Los estudios colaborativos, en general, utilizan sistemas electrónicos de introducción de los datos con ordenadores a través de Internet.

También es necesario establecer un sistema de gobierno con un comité de dirección formado por los IP y representantes de la institución financiadora, y con varios subcomités. Un subcomité debe ser responsable de los aspectos de control de calidad, desarrollando los procedimientos de normalización y los sistemas para la formación, la certificación y la revisión del rendimiento del personal del estudio. Estos tienden a ser complicados y caros, y precisan una formación centralizada para el personal destacado de cada centro, visitas en el centro para la revisión del rendimiento, y auditorías de los datos por parte de personal del centro coordinador y revisores externos (v. apéndice 17B). En otros subcomités generalmente hay grupos que supervisan la inclusión y las actividades clínicas, un grupo que revisa y autoriza las publicaciones y las presentaciones, y otro que evalúa los estudios auxiliares propuestos.

En un estudio multicéntrico, los cambios en las definiciones operativas y otros métodos de estudio surgen, con frecuencia, de preguntas emitidas por un centro clínico y que responde el correspondiente personal o comité del estudio, y que se publican en la página web del estudio en una lista actualizada para que todo el que participa en el estudio conozca los cambios. Si se acumula un número significativo de cambios, deben prepararse las páginas revisadas del manual de instrucciones y de otros documentos del estudio para incluir estos cambios. En los estudios pequeños realizados en un solo lugar se puede seguir un patrón más sencillo, con notas sobre los cambios que se fechan y se conservan en el manual de instrucciones.

Una última idea

Un error frecuente en la investigación es la tendencia a recopilar demasiados datos. El hecho de que el período inicial es la única

posibilidad de medir las variables basales lleva al deseo de incluir todo lo que pueda parecer de interés, y hay una tendencia a realizar más consultas de seguimiento y a recopilar más datos en ellas de lo que puede ser útil. Los investigadores tienden a recopilar muchos más datos de los que nunca analizarán o publicarán.

Un problema que se plantea con ello es el tiempo y los costes necesarios para medir cosas de menor importancia; los participantes se cansan y se aburren, y se deteriora la calidad de las determinaciones más importantes. Otro problema es el tamaño y la complejidad añadidos de la base de datos, lo que dificulta el control de calidad y el análisis de los datos. Es prudente preguntarse la necesidad de cada variable que se recogerá, y eliminar muchas de las que son opcionales. La inclusión de algunas redundancias intencionadas puede mejorar la validez de variables importantes, pero la parquedad es la norma general.

Volver al principio

▪ RESUMEN

- El éxito de la realización del estudio empieza con la reunión de los recursos, entre ellos el espacio, el personal y la financiación del estudio y su puesta en marcha, todo lo cual precisa un sólido liderazgo del IP.
- La puesta en marcha precisa la gestión del presupuesto, la obtención de la autorización del CEIC y la finalización del protocolo y el manual de instrucciones mediante un proceso de estudio preliminar de la adecuación y la viabilidad de los planes de la inclusión, las intervenciones, las mediciones de las variables predictivas y de respuesta, los formularios y la base de datos; el objetivo es minimizar la necesidad de revisiones posteriores del protocolo una vez que haya comenzado la obtención de los datos.
- Las revisiones menores del protocolo una vez comenzado el estudio, como añadir una pregunta a un cuestionario o modificar una definición operativa, son relativamente fáciles de realizar, aunque a veces se precise la aprobación del CEIC y pueda afectarse el análisis de los datos.
- Las revisiones mayores del protocolo una vez comenzado el estudio, como un cambio en la naturaleza de la intervención, los criterios de inclusión o la variable de respuesta primaria, tienen implicaciones importantes, y deben realizarse con renuencia y con la aprobación de organismos esenciales, como el CSDS, el CEIC y la entidad financiadora.
- Hacen falta procedimientos de cierre para informar adecuadamente a los participantes de los hallazgos del estudio y gestionar la transición y las implicaciones de su asistencia.
- Se debe garantizar el control de calidad durante el estudio con un método sistemático bajo la supervisión de un coordinador del control de calidad, siguiendo los principios de las prácticas clínicas correctas (PCC) e incluyendo:
    - Procedimientos operativos estándar (POE) con un manual de instrucciones; formación y certificación del personal, y revisión del rendimiento; informes periódicos (sobre inclusión de participantes, cumplimiento de las consultas y las mediciones); y reuniones frecuentes del equipo.
    - Control de calidad de los procedimientos de laboratorio: enmascaramiento y etiquetado sistemático de las muestras obtenidas de los participantes en el estudio, y uso de mezclas estándar, duplicados con enmascaramiento y mediciones de consenso.
    - Control de calidad de la gestión de los datos: diseño de formularios y sistemas electrónicos para facilitar la supervisión de la exhaustividad, la exactitud y la integridad de la recogida, la introducción, la edición y el análisis de los datos.
- Los estudios colaborativos multicéntricos crean subcomités y otros sistemas distribuidos para la gestión del estudio y el control de calidad.

Volver al principio

[1]Nota: Este es un modelo para un gran ensayo multicéntrico. El manual de instrucciones para un estudio pequeño puede ser menos elaborado.

[2]Las tablas deben contener resultados del período completo del estudio y, cuando proceda, del período desde la redacción del último informe. Cuando sea necesario, deben proporcionarse tasas y comparaciones entre el personal y las unidades participantes.

APÉNDICE 17AEjemplo de una tabla de contenidos de un manual de instrucciones [1]
NA

Capítulo 1. Protocolo del estudio

Capítulo 2. Organizaciones y políticas

Unidades participantes (centros clínicos, laboratorios, centro de coordinación, etc.) e investigadores y personal

Administración y gobierno (comités, entidad financiadora, supervisión de los datos y la seguridad, etc.)

Directrices sobre políticas (publicaciones y presentaciones, estudios auxiliares, conflictos de intereses, etc.)

Capítulo 3. Inclusión de los participantes

Elegibilidad y criterios de exclusión

Diseño del muestreo

Métodos de inclusión de participantes (publicidad, contactos de derivación, cribado, etc.)

Consentimiento informado

Capítulo 4. Consultas clínicas

Contenido de la primera consulta

Contenido y duración de las consultas de seguimiento

Procedimientos de seguimiento de los participantes que no responden

Capítulo 5. Procedimientos de aleatorización y enmascaramiento

Capítulo 6. Variables predictivas

Procedimientos de medición

Intervención, incluyendo los procesos de etiquetado de los fármacos, distribución y manipulación

Evaluación del cumplimiento

Capítulo 7. Variables de respuesta

Valoración y adjudicación de las respuestas primarias

Valoración y gestión de otras respuestas y acontecimientos adversos

Capítulo 8. Control de calidad

Generalidades y responsabilidades

Formación en los procedimientos

Certificación del personal

Mantenimiento del equipo

Revisión externa y visitas al centro

Informes periódicos

Capítulo 9. Gestión de los datos

Recogida y registro de datos

Introducción de datos

Edición, almacenamiento y copias de seguridad

Confidencialidad

Capítulo 10. Planes de análisis de los datos

Capítulo 11. Directrices para la supervisión de los datos y la seguridad

Apéndices

Cartas a los participantes, proveedores principales, etc.

Cuestionarios, formularios

Detalles sobre procedimientos, criterios, etc.

Materiales para la inclusión (anuncios, folletos, cartas, etc.)

APÉNDICE 17BTablas y listas de verificación del control de calidad
NA

- Tabulaciones para supervisar las características del rendimiento [2]
    - Características clínicas
        - Inclusión
            - Número de participantes detectados para su inclusión; número de participantes excluidos y tabulación de las razones para el rechazo
            - Gráfica acumulada del número incluido, en comparación con el que es necesario para lograr el objetivo de inclusión
        - Seguimiento
            - Número de exploraciones de seguimiento realizadas para cada consulta esperada; número de participantes atendidos en un marco temporal especificado
            - Mediciones del cumplimiento de la intervención del estudio, las visitas y las mediciones
            - Número de abandonos y participantes a los que no puede localizarse para el seguimiento
        - Cantidad y calidad de los datos
            - Número de formularos cumplimentados, número que generó mensajes de edición, número de solicitudes de consulta no respondidas, tiempo hasta la resolución de las consultas
            - Número de formularios ausentes, número o proporción de variables ausentes
        - Cumplimiento del protocolo
            - Número de participantes no elegibles incluidos
            - Resumen de datos del recuento de comprimidos y otras medidas de cumplimiento por el grupo de tratamiento
    - Características del centro de datos
        - Número de formularios recibidos y número de entradas de datos en espera

        - Lista acumulativa de cambios de codificación y del protocolo
        - Calendario que indica las tareas realizadas y no acabadas
    - Características del laboratorio central
        - Número de muestras recibidas y número de muestras analizadas
        - Número de muestras identificadas inadecuadamente, perdidas o destruidas
        - Número de muestras que precisan nuevo análisis, y tabulación de los motivos
        - Media y varianza de las diferencias duplicadas con enmascaramiento, y análisis de la tendencia secular, basados en determinaciones repetidas de patrones conocidos
    - Características del centro de lectura
        - Número de registros recibidos y leídos
        - Número de registros recibidos que se etiquetaron incorrectamente o que presentaban otras deficiencias (tabular las deficiencias)
        - Análisis de lecturas repetidas como comprobación de la reproducibilidad de las lecturas y como para controlar las desviaciones temporales en el proceso de lectura
- Componentes de la visita:
    - Visita al centro clínico
        - Reunión privada de los visitantes al centro con el IP
        - Reunión de los visitantes al centro con miembros del equipo médico
        - Inspección de las instalaciones en las que se realizan las exploraciones y del almacén de los registros
        - Comparación de los datos contenidos en formularios de datos seleccionados aleatoriamente con los que contiene el archivo de datos informático
        - Revisión del archivo de formularios de datos y registros relacionados para valorar la exhaustividad y la seguridad frente a la pérdida o el mal uso
        - Observación del personal clínico cuando realiza procedimientos especificados
        - Comprobación de manuales de instrucciones, formularios y otros documentos del archivo del centro para evaluar si están actualizados
        - Observación o descripción verbal de determinados procedimientos (p. ej., la serie de exploraciones necesarias para determinar la elegibilidad de los participantes)
        - Conversaciones con participantes reales del estudio durante su participación o tras ella, como comprobación del proceso de consentimiento informado
        - Conversaciones privadas con personal de apoyo clave para valorar sus prácticas y su filosofía con respecto a la recogida de los datos
        - Reuniones privadas con el IP acerca de los problemas identificados
    - Visita al centro de datos
        - Revisión de los métodos para inventariar los datos recibidos de las consultas
        - Revisión de los métodos para la gestión y la comprobación de los datos

- Valoración de la idoneidad de los métodos para archivar y almacenar los registros en papel recibidos de las consultas, incluyendo la seguridad de la zona de almacenamiento y los métodos para proteger los registros frente a la pérdida o el uso no autorizado
- Revisión de los recursos informáticos disponibles
- Revisión del método de aleatorización y de las garantías para proteger frente a las alteraciones del proceso de aleatorización
- Revisión de los procedimientos de edición de los datos y seguimientos de auditoría
- Revisión de la estructura de los archivos de datos informáticos y métodos para mantener la base de datos de análisis
- Revisión de los métodos de programación para la gestión y el análisis de datos, incluyendo una valoración de la documentación del programa
- Comparación de la información contenida en los formularios del estudio original con los del archivo de datos informático
- Revisión de los métodos para generar archivos del análisis de los datos y los correspondientes informes de los datos
- Revisión de los métodos para hacer copias de seguridad del archivo de datos principal
- Revisión del archivo maestro de los documentos esenciales del estudio, como manuales, guías, formularios de datos, actas de los comités del estudio, etc., para determinar que están completos

BIBLIOGRAFÍA

1. Mosca L, Barrett-Connor E, Wenger NK, et al. Design and methods of the Raloxifene Use for The Heart (RUTH) Study. Am J Cardiol 2001;88:392-395. Citado aquí

2. MORE Investigators. The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial. Multiple outcomes of raloxifene evaluation. JAMA 1999;281:2189-2197. Citado aquí

3. Shepherd R, Macer JL, Grady D. Planning for closeout-From day one. Contemp Clin Trials 2008;29:136-139 Citado aquí

4. http://www.fda.gov/downloads/Drugs/Guidances/ucm073122.pdf Citado aquí

5. FDA Regulations Relating to Good Clinical Practice and Clinical Trials. Available at: www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/ucm114928.htm Citado aquí

6. Information about Good Clinical Practices in the European Medicines Agency International Conference on Harmonization. Available at: http://www.ich.org or at http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000035.jsp&murl=menus/regulations/regulations.jsp&mid=WC0b01ac0580027645&jsenab Citado aquí

CAPÍTULO 18 Estudios comunitarios e internacionales

Norman Hearst
Thomas Novotny

La mayor parte de la investigación clínica tiene lugar en centros médicos universitarios u otras instituciones académicas extensas, centros que ofrecen muchas ventajas para realizar una investigación, entre ellas la evidente de contar con investigadores con experiencia. Una cultura, una reputación y una infraestructura ya establecidas facilitan el trabajo tanto al investigador principiante como al catedrático con un trabajo fijo. El éxito engendra más éxito, de manera que la investigación clínica se concentra en centros de excelencia. Este capítulo, por el contrario, aborda la investigación que se realiza fuera de estos centros.

Se define la investigación comunitaria como la que tiene lugar fuera del entorno habitual de los centros médicos y tiene como objetivo satisfacer las necesidades de las comunidades en las que se realiza. La investigación internacional, sobre todo en países pobres, puede conllevar muchos de los mismos retos de responder a las necesidades locales y establecer una infraestructura de investigación donde no existía previamente. Además, esta investigación exige el conocimiento de numerosos aspectos políticos, burocráticos y culturales complejos que se originan en la investigación internacional. Ambos tipos de investigación suponen que exista una colaboración entre investigadores locales y compañeros de un centro de investigación establecido. Esa colaboración es fundamental para resolver problemas sanitarios mundiales o locales de larga evolución o nuevos, y pueden ser oportunidades extraordinarias para el crecimiento personal y el aprendizaje mutuo. Sin embargo, estas colaboraciones puede ser difíciles debido a las distancias físicas que separan a los investigadores, las diferencias culturales que afectan a los participantes, los problemas políticos que influyen en las instituciones locales y nacionales, y las limitaciones económicas, tanto en los países donantes como en los receptores.

▪ ¿POR QUÉ UNA INVESTIGACIÓN COMUNITARIA E INTERNACIONAL?

La investigación colaborativa es muchas veces la única forma de abordar preguntas de la investigación que se relacionan con contextos especiales, enfermedades nuevas o que están reapareciendo, y poblaciones específicas. La investigación en centros médicos académicos tiende a centrarse en prioridades de ciencia clínica o básica que pueden diferir bastante de las necesidades de la comunidad local, y pueden ser aún más diferentes de los problemas sanitarios mundiales que afectan a grandes segmentos de la población mundial. Esos problemas mundiales precisan esfuerzos colectivos por razones humanitarias y porque las fronteras nacionales, estatales o locales no aíslan las comunidades de los efectos de dichos problemas. La «brecha 10/90» en investigación sanitaria, según la cual el 9% del volumen mundial de enfermedad solo recibe el 1 % de la inversión de investigación mundial (1), es una sólida justificación para que se realice más investigación colaborativa dirigida a los enormes problemas sanitarios de los países de renta baja o media (PRBM). Por lo tanto, existe la necesidad de reforzar la capacidad investigadora de los PRBM y de las comunidades mediante colaboraciones internacionales y nacionales. Esto incluye una cuidadosa atención al desarrollo de procesos de revisión institucional y a la protección de los participantes humanos. Además, la participación en el proceso de la investigación ofrece a una comunidad y a los investigadores unos beneficios que van más allá del valor de la información recopilada en un estudio concreto. La investigación comunitaria que se realiza con atención y preocupación por el bien público genera relaciones duraderas, un sentido de orgullo y. tal vez. incluso el desarrollo económico.

Preguntas locales

Muchas preguntas de investigación precisan respuestas que solo se consiguen mediante la investigación local. Los datos a nivel estatal o nacional a partir de fuentes centrales pueden no reflejar exactamente la magnitud de las enfermedades en el ámbito local ni la distribución de los factores de riesgo en la comunidad local. Las intervenciones, especialmente las que se dirigen a modificar la conducta, pueden no tener el mismo efecto en diferentes entornos. Por ejemplo, la eficacia en la salud pública de la promoción del preservativo como estrategia de prevención del VIH/sida es bastante diferente en Estados Unidos y en África (2). Para hallar métodos que encajen en las necesidades locales, se precisa la investigación local (tabla 18-1).

Los datos biológicos sobre la fisiopatología de las enfermedades y la eficacia de los tratamientos suelen poder generalizarse a una gran variedad de poblaciones y culturas. Sin embargo, puede haber diferencias raciales, culturales o genéticas, o diferencias basadas en las causas regionales de las enfermedades, que hacen que sea necesaria una investigación local. Por ejemplo, la eficacia de los fármacos antihipertensivos es diferente en pacientes descendientes de africanos y europeos (3); los microorganismos causales y los patrones de sensibilidad antimicrobiana en la neumonía son diferentes en Bolivia y en Boston; y la percepción de la salud, la asistencia sanitaria y la enfermedad pueden diferir mucho de unas comunidades a otras (4).

Mayor posibilidad de generalización

La investigación comunitaria es a veces útil para producir resultados que son más generalizables. Por ejemplo, los pacientes con dolor lumbar que han sido atendidos en hospitales de referencia son muy diferentes de los pacientes que acuden con dolor lumbar a los médicos de atención primaria. Los estudios de la evolución natural del dolor lumbar o la respuesta al tratamiento en un centro de

asistencia terciaria, por lo tanto, pueden tener un uso limitado para la práctica clínica en la comunidad.

En parte como respuesta a este problema, se han organizado varias redes de investigación basada en la práctica, en las que los médicos de entornos comunitarios trabajan juntos para estudiar preguntas de investigación de interés mutuo (5). Un ejemplo es la respuesta al tratamiento de pacientes con síndrome del túnel carpiano en las consultas de atención primaria (6). La mayoría de los pacientes mejoró con un tratamiento conservador; algunos precisaron una consulta con un especialista o sofisticadas pruebas diagnósticas. En estudios previos se había recomendado la intervención quirúrgica temprana en el síndrome del túnel carpiano, de acuerdo con estudios de pacientes tratados en un importante centro de referencia.

La posibilidad de generalización también es importante en la investigación internacional. Los hallazgos de la investigación en un país no siempre pueden aplicarse en otro. Pero, aunque los resultados se generalizan mejor al lugar donde se realizó la investigación, también pueden ser importantes para poblaciones que migran y cuyo origen está en el país de la investigación. Estas poblaciones que se desplazan son cada vez más importantes en un mundo globalizado que tuvo 175 millones de emigrantes internacionales en 2010 (7). La globalización obliga, actualmente, a tener una perspectiva más amplia sobre el riesgo de enfermedad, y también sobre los abordajes de investigación colaborativa para acordar enfermedades que atraviesan con tanta facilidad las fronteras nacionales.

Generación de capacidad en el ámbito local

La investigación clínica no debe ser propiedad exclusiva de los centros médicos universitarios. Las prioridades de los investigadores de estos centros deben reflejar las prioridades de las agencias que financian la investigación, los problemas con los que se encuentran en su práctica diaria, y lo que creen que tienen importancia científica o económica general. La realización de investigaciones en la comunidad e internacional garantiza que también se prioricen cuestiones de importancia local (8).

¿Cuál es la frecuencia de uso del asiento infantil y el cinturón de seguridad en una vecindad de renta baja de Chicago?

¿Cuáles son los patrones de resistencia a antimicrobianos de cepas de tuberculosis en Uganda?

¿Cuál es el efecto de una campaña de prevención de las ETS en el lugar de trabajo para granjeros emigrantes en Texas?

¿Qué proporción de cardiopatía isquémica en mujeres de Brasil se asocia al tabaquismo?

El valor de la participación de la comunidad en investigación va más allá de la información específica recopilada en cada estudio. Investigar tiene un triple efecto muy positivo, al elevar los patrones de conocimientos locales y fomentar la creatividad y el pensamiento independiente. Cada proyecto supone la adquisición de habilidades y confianza, que permiten que los investigadores locales se contemplen como participantes plenos en el progreso científico y no solo como consumidores de conocimientos producidos en otros lugares. Esto, a su vez, fomenta más investigación. Además, participar en una investigación puede proporcionar recursos intelectuales y económicos a una comunidad, y ayuda a fomentar la capacitación y la autosuficiencia locales.

Volver al principio

▪ INVESTIGACIÓN COMUNITARIA

En teoría, la puesta en marcha de la investigación comunitaria es el mismo proceso que para cualquier otra tarea de investigación. El método general esbozado en este libro se aplica igual en una pequeña ciudad de Estados Unidos que en Nepal, San Francisco o Londres. En la práctica, el mayor reto es encontrar compañeros con experiencia o mentores con quienes trabajar y aprender. Esta ayuda puede no estar disponible a nivel local, lo que a menudo hace que los investigadores comunitarios e internacionales tengan que tomar una importante decisión temprana: trabajar solos o en colaboración con investigadores con más experiencia que se encuentran en otros lugares.

Empezar sin compañía

Empezar en la investigación sin la ayuda de un compañero con más experiencia es como aprender a nadar solo: no es imposible, pero sí difícil, y en ocasiones acechan peligros imprevistos. Sin embargo, muchas veces es la única opción. Seguir unas pautas puede hacer que

el proceso sea algo más fácil.

- Comience por lo sencillo. Rara vez es una buena idea empezar a investigar en una comunidad con un ensayo aleatorizado y controlado. Los pequeños estudios piloto descriptivos que generan datos locales útiles pueden tener más sentido: es mejor conseguir un pequeño éxito que un gran fracaso. Los proyectos más ambiciosos deben guardarse para más adelante, y se pueden aprovechar los datos piloto que se han generado previamente. Por ejemplo, un estudio descriptivo sobre el uso de preservativos en varones jóvenes en Uganda realizado por un investigador local principiante sirvió como primer paso para un mayor ensayo de intervención sobre la prevención del VIH/sida en esa comunidad (9).

- Piense en las ventajas comparativas locales. ¿Qué preguntas puede responder un investigador en su entorno local mejor que en ningún otro lugar? Esto suele significar dejar el desarrollo de nuevas técnicas analíticas y nuevos tratamientos a los centros médicos universitarios y a las organizaciones de investigación farmacéutica. Con frecuencia es mejor que un investigador joven se centre en problemas sanitarios o en poblaciones que no suelen encontrarse en otros lugares, pero que sí son habituales en su comunidad.

- Red. Como se comentó en el capítulo 2, trabajar en red es importante para todo investigador. Un nuevo investigador debe establecer todos los contactos que pueda con científicos de todas partes que estén tratando cuestiones similares. Si no dispone de colaboradores formales, puede que al menos sea posible encontrar a alguien con quien comentar un borrador de un protocolo de investigación, un cuestionario o un manuscrito mediante correo electrónico o por teléfono. Asistir a una conferencia científica sobre su campo de interés es una buena forma de establecer estos contactos, y la remisión al trabajo de un compañero con mayor experiencia puede ser una buena forma de iniciar dicho contacto.

Investigación colaborativa

Ya que es difícil empezar uno solo, un buen modo de comenzar a investigar en una comunidad suele ser haciéndolo en colaboración con investigadores más experimentados de otros lugares, especialmente si esos investigadores ya disponen de confianza, contactos y metodologías en la población que se va a estudiar. Hay dos modelos principales para esta colaboración: descendente y ascendente (10).

El modelo descendente se refiere a estudios que se originan en un centro universitario y en los que investigadores de la comunidad intervienen en la recopilación de participantes y en la realización del estudio. Esto sucede, por ejemplo, en grandes ensayos multicéntricos que invitan a consultas y a hospitales a incluir pacientes en un protocolo de investigación ya establecido. Este método tiene la gran ventaja de que cuenta con colaboradores con experiencia que suelen responsabilizarse del diseño del estudio y de obtener los recursos y las autorizaciones necesarias para llevarlo a cabo.

En el modelo ascendente, los investigadores con experiencia orientan y proporcionan asistencia técnica a investigadores locales y comunidades para que desarrollen sus propias agendas de investigación. Algunos centros médicos universitarios ofrecen programas de formación para investigadores de la comunidad o internacionales. Si se logra acceder a un programa de este tipo o se consigue establecer una relación equivalente, puede ser magnífico para desarrollar la capacidad de investigación local, especialmente cuando esta relación se mantiene a largo plazo. Sin embargo, establecer una relación de este tipo no es fácil. La mayor parte de las entidades financiadoras están más interesadas en patrocinar proyectos de investigación específicos que en dedicar recursos a generar la capacidad de investigación local y colaboraciones de investigación. Incluso aunque se disponga de fondos para cubrir los gastos de formación y viaje, los investigadores con experiencia pueden preferir dedicar su tiempo a realizar su propia investigación, antes que ayudar a otros a empezar. Pese a ello, nunca se podrá insistir lo suficiente en el valor en cuanto a satisfacción, importancia y relevancia para la comunidad local de la investigación participativa colaborativa de base comunitaria (IPCC), en la que la comunidad participa por completo en todos los aspectos de la investigación (11).

Los investigadores de la comunidad deben aprovechar los posibles incentivos que pueden ofrecer a investigadores con más experiencia con los que les gustaría trabajar. En el modelo descendente, lo más importante que pueden ofrecer es el acceso a participantes. En el modelo ascendente, los incentivos pueden ser el mérito científico intrínseco de un estudio realizado en la comunidad, la autoría compartida de las publicaciones resultantes, y la satisfacción de elaborar una relación de colaboración y de ayudar a una comunidad a desarrollar su capacidad investigadora.

Para empezar un nuevo programa de investigación, la opción ideal puede ser establecer una relación a largo plazo con una institución de investigación ya establecida. Las agencias colaboradoras pueden firmar memorandos de entendimiento (ME) para que se puedan ofrecer a los posibles financiadores datos por escrito de la comunicación y los acuerdos. Haber establecido por anticipado esta colaboración puede ahorrar tiempo y frustración. La colaboración de este tipo puede conllevar una combinación de proyectos descendentes y ascendentes. Hay que recordar, no obstante, que la buena colaboración es, fundamentalmente, entre investigadores individuales. Una institución académica puede proporcionar el clima, la estructura y los recursos que fomenten la colaboración individual, pero son las personas las que proporcionan la sensibilidad cultural, el respeto mutuo, el trabajo duro y el compromiso a largo plazo para que todo

funcione.

Volver al principio

▪ INVESTIGACIÓN INTERNACIONAL

La investigación internacional conlleva, con frecuencia, la colaboración entre grupos con niveles diferentes de experiencia y recursos, y por ello está sujeta a muchos de los mismos problemas que la investigación comunitaria. Sin embargo, la investigación internacional plantea otros retos. Los problemas que se describen en el apartado siguiente son especialmente importantes.

Barreras de distancia, idioma y cultura

Sin un conocimiento completo de las perspectivas culturales de la comunidad, muchos investigadores encuentran que incluso los planes mejor organizados fracasan a pesar de una planificación cuidadosa y de utilizar tecnologías avanzadas. Para evitar el fracaso, los investigadores deben conocer las percepciones culturales de la enfermedad en las comunidades en las que pretenden trabajar, y deben desarrollar abordajes sólidos desde el punto de vista cultural para su investigación en colaboración. Debido a las distancias, las oportunidades para la comunicación cara a cara entre compañeros internacionales son limitadas. Si es posible, los colaboradores deben realizar al menos una visita al centro del otro. Las conferencias internacionales pueden proporcionar, a veces, oportunidades para reunirse, pero probablemente sean pocas. Afortunadamente, el correo electrónico, Internet y Skype han hecho que las comunicaciones internacionales sean más fáciles, rápidas y económicas. Se puede tener una buena comunicación a cualquier distancia, pero ambas partes deben esforzarse. Los métodos más modernos no sirven de nada si no se usan con frecuencia. La falta de comunicación frecuente y de respuesta rápida a las consultas, por una u otra parte, es un signo de que la colaboración puede tener problemas.

Las diferencias de idioma se superponen a menudo a las barreras de comunicación causadas por la distancia. Si el primer idioma hablado por los investigadores no es el mismo en todos los lugares, es importante que haya uno que todo el mundo pueda usar (habitualmente ese idioma es el inglés). Suponer que todas las interacciones sean en inglés, sin embargo, coloca a los investigadores de algunos países en desventaja. Es poco probable que investigadores extranjeros que no hablen el idioma local tengan más que un conocimiento superficial de la cultura del país, y no podrán participar totalmente en muchos aspectos esenciales de un estudio, como el desarrollo y la validación de un cuestionario. Esta comunicación tiene mucha importancia en estudios con componentes conductuales.

Aunque se superen las barreras lingüísticas, las diferencias culturales pueden causar graves malentendidos entre los investigadores y los participantes, o entre los propios investigadores. Las traducciones literales de cuestionarios pueden tener significados diferentes, ser culturalmente inadecuadas u omitir factores locales esenciales. Las normas institucionales pueden ser diferentes. Por ejemplo, en algunos marcos, un jefe de departamento de colaboradores extranjeros con escasa implicación directa en un estudio podría esperar ser el primer autor de la publicación resultante. Estos temas deben tratarse y aclararse por adelantado, como parte del proceso de desarrollo institucional del proyecto. La paciencia, la buena voluntad y la flexibilidad suelen solucionar los problemas de este tipo. En proyectos de mayor envergadura puede ser aconsejable incluir un antropólogo, un ético u otro experto en temas culturales como parte del equipo de investigación.

Es esencial que exista una comunicación frecuente, clara y fluida, así como la rápida aclaración de cualquier duda o confusión. Al tratar con diferencias culturales e idiomáticas, es mejor ser repetitivo y arriesgarse a insistir en lo evidente que hacer suposiciones incorrectas sobre lo que piensan o dicen los demás. Los acuerdos por escrito que explican las obligaciones y responsabilidades mutuas pueden ayudar a aclarar puntos como la propiedad de los datos, el orden de los autores, los derechos sobre la publicación y las decisiones sobre la estructura de los resultados de la investigación. El desarrollo de estos acuerdos precisa la atención personal y cuidadosa de los colaboradores de ambas partes.

Asuntos de financiación

Debido a las desigualdades económicas, la colaboración entre instituciones de países pobres y ricos solo suele ser posible con financiación por parte del país con más recursos o, con menos frecuencia, de otros países u organizaciones internacionales. Cada vez hay más organizaciones activas en la investigación sanitaria mundial, pero su apoyo a menudo está limitado por una agenda específica con requisitos estrictos de resultados medibles. Buena parte de la financiación por donantes bilaterales tiende a fluir a través de la institución del país rico, lo que refuerza la posición subordinada de las instituciones de los PRBM. Como en cualquier situación con un desequilibrio de poder, esto crea dificultades éticas. Cuando los investigadores de los países ricos controlan los recursos económicos, no es infrecuente que traten a sus homólogos de los países menos favorecidos más como a empleados que como a compañeros. Las entidades financiadoras y los donantes internacionales deben tener especial cuidado en evitar esto y fomentar, en su lugar, una dirección conjunta real de las actividades colaborativas (8).

Las diferentes prácticas de gestión económica constituyen otra posible fuente de conflictos entre los miembros del consorcio de investigación. Las instituciones de los países ricos pueden tratar de imponer patrones de contabilidad difíciles o imposibles de cumplir localmente. Las instituciones de los PRBM pueden cargar los presupuestos con ordenadores y otros equipos que esperan conservar tras finalizar el estudio. Aunque esto es comprensible a la vista de sus necesidades y de la ausencia de fuentes de financiación alternativas, es importante que todos los subsidios que vayan más allá del coste real de la realización de la investigación se negocien claramente, y que se apliquen prácticas de contabilidad para satisfacer las necesidades de las agencias financiadoras. Por el contrario, los gastos generales institucionales y los salarios de los investigadores elevados a menudo crean una situación injusta en la que la mayor parte de los fondos para las investigaciones colaborativas permanecen en los países donantes, aun cuando la mayor parte del trabajo se realiza en el país socio.

Las instituciones y los donantes de los países donantes deben prestar especial atención a la generación de la capacidad de administración de la investigación de los compañeros locales. Esto puede suponer la formación administrativa y presupuestaria o el uso de especialistas en el campo para ayudar en las tareas administrativas locales. Un requisito para los socios internacionales es obtener un número D-U-N-S, un número de identificación exclusivo con nueve dígitos para la localización física de cada una de las instituciones que soliciten contratos o becas del gobierno federal estadounidense (http://fedgov.dnb.com/webform). El esfuerzo invertido en desarrollar unos medios administrativos puede redundar en una mejor respuesta a los plazos, una comunicación más eficaz, evitar conflictos innecesarios y construir una infraestructura sólida para investigaciones futuras.

Cuestiones éticas

La investigación internacional conlleva cuestiones éticas que deben afrontarse con firmeza. Se aplican aquí todas las cuestiones éticas de la investigación (cap. 4). Debido a que la investigación internacional puede plantear riesgos particulares en relación con la violación de la protección de los participantes humanos, también se necesitan unas garantías y consideraciones adicionales.

¿Cuál es, por ejemplo, el grupo de comparación adecuado cuando se estudian nuevos tratamientos en un PRBM en el que no se dispone del tratamiento convencional? Los testigos con placebo no son éticos cuando hay otros fármacos eficaces que constituyen el tratamiento habitual en otras partes. Pero ¿cuál es el «tratamiento de referencia» en una comunidad en la que la mayoría de las personas es demasiado pobre para permitirse los tratamientos que puedan estar disponibles en muchos países? Por un lado, puede que no sea posible que los investigadores proporcionen el tratamiento más avanzado a todos los participantes de un estudio. Por otro lado, permitir los testigos con placebo simplemente porque hay un acceso inadecuado a los fármacos y la asistencia sanitaria no es ético y ha sido puesto en entredicho por muchos grupos intergubernamentales y organizaciones de defensa de los pacientes. Por ejemplo, los estudios de tratamientos antirretrovíricos orales menos costosos para prevenir la transmisión del VIH de la madre al hijo que se han realizado en países en los que la mayoría de las mujeres no tenían acceso a un régimen terapéutico existente de eficacia demostrada confirman algunos de estos problemas (12 , 13).

Hay un tema relacionado que tiene que ver con el estudio de tratamientos que no es probable que sean económicamente accesibles para la población del país. ¿Son éticos estos estudios, aunque sigan todas las normas habituales? Por ejemplo, ¿sería ético estudiar un nuevo fármaco para la diabetes del tipo II en un PRBM en el que probablemente los pacientes no pudieran permitirse comprar este fármaco? Son preguntas que no tienen respuestas sencillas. Se han puesto objeciones a las convenciones internacionales establecidas que regulan la investigación ética, como la Declaración de Helsinki, y están sujetas a múltiples interpretaciones (14 , 15).

Una prueba clave sería considerar, en primer lugar, por qué se está realizando el estudio en un PRBM. Si el objetivo principal es conseguir información para ayudar a las personas de ese país, esto iría a favor del estudio y se debería considerar en consecuencia. De modo ideal, el objetivo de la investigación debe ser un cambio sostenible y un valor añadido para el país (16). Si, por el contrario, el objetivo es la conveniencia o evitar los obstáculos de realizar el estudio en un país rico, el estudio debe someterse a todos los requisitos éticos que se aplicarían en el país que lo patrocina, incluyendo el importante requisito de la justicia distributiva (cap. 14).

Por estos y otros motivos, los estudios en países sin recursos económicos que están dirigidos o financiados desde algún otro lugar deben ser aprobados por comités éticos de ambos países. Pero, aunque es necesaria esta aprobación, no garantiza que un estudio sea ético. Los sistemas de revisión ética de la investigación en muchos países sin recursos son débiles o inexistentes, y pueden, a veces, estar manipulados por políticos o investigadores locales. Por el contrario, los comités de revisión de los países ricos son a veces ignorantes o insensibles ante los temas especiales que implica la investigación internacional. La aprobación oficial no elimina la responsabilidad final de la realización ética de la investigación por parte de los propios investigadores.

Otro importante problema ético es el tratamiento de los colaboradores de los PRBM socios. Varios son los puntos que deben acordarse de antemano. ¿A quién pertenecen los datos que se generarán? ¿Quién necesita la autorización de quién para realizar y publicar los análisis? ¿Tendrán los investigadores locales el apoyo que precisan para preparar manuscritos para la publicación internacional sin tener que pagar por ello dejando de ser los primeros autores? ¿Cuánto dura un compromiso por ambas partes? Un ensayo de gran tamaño, realizado en varios países con escasos recursos, sobre asesoramiento y pruebas voluntarias para evitar la infección por el VIH, excluyó

de repente su centro de colaboración en Indonesia (17). Según los investigadores, se debió a que la variable de respuesta de interés se hizo menos frecuente en ese lugar de lo que se había proyectado en los cálculos de potencia del estudio. Aun cuando esta decisión tenía sentido práctico, los indonesios lo consideraron un engaño.

Hay otras cuestiones éticas que pueden tener que ver con realidades políticas y económicas locales. Por ejemplo, un ensayo clínico planificado de la profilaxis preexposición frente al VIH con tenofovir para trabajadores del sexo se suspendió, aun cuando había sido aprobado por comités éticos multinacionales (18). A los posibles participantes del estudio les preocupó que pudieran no contar con asistencia médica por problemas relacionados con la infección por el VIH o con los efectos de los fármacos, y no estaban dispuestos a participar sin garantías de seguros médicos de por vida. El primer ministro del país intervino para interrumpir el ensayo.

Finalmente, un objetivo explícito de toda colaboración internacional debe ser aumentar la capacidad de investigación local. ¿Qué habilidades y equipamiento dejará el proyecto tras de sí cuando finalice? ¿Qué actividades se realizarán para formar al personal del proyecto? ¿Participarán los investigadores locales en conferencias internacionales? ¿Participarán tan solo los investigadores locales de alto nivel que ya han tenido muchas oportunidades, o tendrán también alguna posibilidad los compañeros más jóvenes? ¿Serán los investigadores locales verdaderos colaboradores y autores principales de la publicación, o simplemente se cuenta con ellos para recopilar los datos? Los científicos de los países con escasos recursos deben preguntar y esperar respuestas claras a estas preguntas. Como se resume en la tabla 18-2 , una buena comunicación y un compromiso a largo plazo son temas recurrentes en la investigación colaborativa internacional de éxito.

La Organización Mundial de la Salud ha publicado recientemente un conjunto de casos en los que se abordan problemas éticos en la investigación sanitaria mundial (19) para ayudar a que los investigadores, los miembros de los comités éticos, las autoridades sanitarias y otras partes implicadas ejerzan sus correspondientes funciones en la realización ética de una investigación. Se puede aprender mucho de los errores y éxitos de los demás, aunque con buena voluntad por parte de los financiadores, los socios de países donantes y los funcionarios de las dos partes de las colaboraciones de la investigación se pueden garantizar los principios éticos en una investigación internacional, y se puede reforzar a nivel mundial la capacidad de realizar dicha investigación.

Científicos de países de ingresos bajos y medios (PRBM)

Elegir colaboradores cuidadosamente

Aprender inglés (u otro idioma de los colaboradores)

Familiarizarse con la bibliografía científica internacional en el área del estudio

Asegurarse de que la colaboración generará la capacidad de investigación local

Dejar claras las expectativas administrativas y científicas de antemano

Científicos de países de ingresos elevados

Elegir colaboradores cuidadosamente

Aprender el idioma y la cultura locales

Comprender los asuntos éticos locales

Fomentar la colaboración local en todos los aspectos del proceso de la investigación

Dejar claras las expectativas administrativas y científicas de antemano

Organismos financiadores

Establecer prioridades de financiación según las necesidades de salud pública

Fomentar la verdadera colaboración, en lugar de un modelo puramente «descendente»

Reconocer la importancia de generar la capacidad de investigación local

Establecer subvenciones para equipo e infraestructura en el ámbito local

Asegurarse de que los gastos generales y los sueldos en el país de rentas elevadas no se apropien de una parte excesiva del presupuesto

Riesgos y frustraciones

Los investigadores de países ricos que piensan en participar en investigaciones internacionales deben empezar por observar de forma realista las dificultades y los riesgos que eso supone. Acometer un trabajo de ese tipo suele ser un proceso largo y lento. Los obstáculos burocráticos son frecuentes en ambas partes. En los países que carecen de infraestructura y estabilidad política, años de trabajo pueden verse afectados por una importante interrupción a causa de catástrofes naturales o causadas por el hombre. En casos extremos, estas situaciones pueden amenazar la seguridad del personal del proyecto o de los investigadores. Por ejemplo, importantes programas colaborativos en la investigación del VIH/sida que habían costado años de trabajo se destruyeron completamente por las guerras civiles en Ruanda y el Congo. Otras dificultades menos catastróficas y más habituales son los problemas cotidianos y los riesgos sanitarios a los que pueden enfrentarse los investigadores expatriados, que van desde las aguas contaminadas y el paludismo hasta la niebla contaminada, la delincuencia común y los accidentes de tráfico.

Otra frustración para los investigadores en PRBM es la dificultad para aplicar sus hallazgos. Incluso aunque pudieran desarrollarse nuevas estrategias para evitar o tratar enfermedades y se demostrara su eficacia, la falta de recursos y de voluntad política impide, a menudo, su aplicación de forma extensa en los países anfitriones. Los investigadores deben ser realistas en sus expectativas, amoldar su trabajo a las estrategias de investigación que podrían realizarse si se demostrara su eficacia, y estar preparados para actuar como defensores para mejorar la salud de la población que estudian.

Recompensas

A pesar de las dificultades, la necesidad de más investigación sanitaria en muchas partes del mundo es abrumadora. Al participar en investigaciones internacionales, un investigador de un país donante puede tener, a veces, un efecto sobre la salud pública mayor y más importante de lo que sería posible si permaneciera entre los muros del entorno académico. El efecto no se debe solo a la propia investigación, sino también a lo que a veces se denomina diplomacia sanitaria mundial. De hecho, actualmente se considera que la salud es una importante fuerza impulsora de las prioridades en política exterior (20). La diplomacia sanitaria se puede llevar a la práctica mediante investigación colaborativa en problemas de salud mundial, como el VIH/sida, el paludismo, la TB, la salud maternoinfantil y el refuerzo de los sistemas sanitarios. La salud y la política siempre han estado interrelacionadas, aunque en un mundo globalizado hay una necesidad creciente de acciones corporativas para abordar los principales problemas sanitarios transfronterizos; la investigación internacional forma parte de este esfuerzo mundial. La probabilidad de tener una implicación significativa y de hacer una contribución real a la salud mundial es un privilegio que puede enriquecer nuestras carreras profesionales y nuestras vidas personales. Todos tenemos que ganar gracias a una mayor colaboración y a una ampliación de las posibilidades de investigación.

Volver al principio

▪ RESUMEN

- La investigación comunitaria e internacional es necesaria para descubrir diferencias regionales en aspectos como la epidemiología de una enfermedad, y los factores culturales y otros factores locales que determinan qué intervenciones serán eficaces.
- La participación local en la investigación clínica puede tener beneficios secundarios para la región, como mejores niveles de conocimiento y autosuficiencia.
- Aunque los problemas teóricos y éticos que conlleva la investigación en estudios comunitarios e internacionales pueden aplicarse ampliamente, cuestiones prácticas, como conseguir fondos y un mentor, son más difíciles en una comunidad; son consejos para el éxito empezar por poco, pensar en ventajas locales y trabajar en red.
- La colaboración entre centros médicos universitarios e investigadores de la comunidad puede seguir un modelo descendente (investigadores de la comunidad realizan estudios que se originan en el centro académico) o un modelo ascendente (investigadores del centro académico ayudan a los investigadores de la comunidad a realizar investigaciones que puedan originar ellos mismos).

- La investigación internacional conlleva muchas de las mismas cuestiones que la investigación comunitaria, con retos adicionales, particularmente en países de ingresos bajos y medios (PRBM), que se relacionan con la comunicación y el idioma, las diferencias culturales, la financiación, el desequilibrio de las estructuras de poder, y las prácticas económicas y administrativas.

- La investigación internacional tiene sus propios problemas éticos, como el estudio de tratamientos que pueden no ser asequibles en los PRBM, el uso de placebos en poblaciones vulnerables, y el estatus y el tratamiento de los colaboradores.

- Superar los retos en la investigación internacional puede proporcionar las recompensas de ayudar a las personas que lo necesitan, formar parte de una comunidad sanitaria mundial mayor y enriquecer las propias experiencias culturales.

Volver al principio

BIBLIOGRAFÍA

1. Unite for Sight. The importance of global health research: closing the 10/90 gap. Available at: http://www.uniteforsight.org/global-impact-lab/global-health-research#_ftnref12, accessed 9/23/12. Citado aquí

2. Hearst N, Chen S. Condom promotion for AIDS prevention in the developing world: is it working? Studies in Family Planning 2004;35(1):39-47. Citado aquí

3. Drugs for hypertension. Med Lett Drugs Ther 1999;41:23-28. Citado aquí

4. Griffith BN, Lovett GD, Pyle DN, et al. Self-rated health in rural Appalachia: health perceptions are incongruent with health status and health behaviors. BMC Public Health 2011;11:229. doi:10.1186/1471-2458-11-229. Citado aquí

5. Nutting PA, Beasley JW, Werner JJ. Practice-based research networks answer primary care questions. JAMA 1999;281:686-688. Citado aquí

6. Miller RS, Ivenson DC, Fried RA, et al. Carpal tunnel syndrome in primary care: a report from ASPN. J Fam Pract 1994;38:337-344. Citado aquí

7. United Nations Department of Economic and Social Affairs (UN DESA). Trends in international migrant stock: the 2008 revision. Available at: http://esa.un.org/migration/index.asp?panel=1, accessed 1/12/2013. Citado aquí

8. Lee K, Mills A. Strengthening governance for global health research: the countries that most need health research should decide what should be funded. BMJ 2009;2000:775-776. Citado aquí

9. Kajubi P, Kamya MR, Kamya S, et al. Increasing condom use without reducing HIV risk: results of a controlled community trial in Uganda. Journal of AIDS 2005;40(1):77-82. Citado aquí

10. Hearst N, Mandel J. A research agenda for AIDS prevention in the developing world. AIDS 1997;11(Suppl 1):S1-4. Citado aquí

11. Minkler M and Wallerstein N, eds. (2008). Community-Based Participatory Research for Health: From Process to Outcomes. ISBN 978-0-470-26043-2. Jossey-Bass Citado aquí

12. 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;337:853-856. Citado aquí

13. Perinatal HIV Intervention Research in Developing Countries Workshop Participants. Science, ethics, and the future of research into maternal-infant transmission of HIV-1. Lancet 1999;353:832-835. Citado aquí

14. Brennan TA. Proposed revisions to the Declaration of Helsinki: will they weaken the ethical principles underlying human research? N Engl J Med 1999;341:527-531. Citado aquí

15. Levine RJ. The need to revise the Declaration of Helsinki. N Engl J Med 1999;341:531-534. Citado aquí

16. Taylor D, Taylor CE. Just and lasting change: when communities own their futures. Baltimore: JHU Press, 2002. Citado aquí

17. Kamenga MC, Sweat MD, De Zoysa I, et al. The voluntary HIV-1 counseling and testing efficacy study: design and methods. AIDS and Behavior 2000;4:5-14. Citado aquí

18. Page-Shafer K, Saphonn V, Sun LP, et al. HIV prevention research in a resource-limited setting: the experience of planning a trial in Cambodia. Lancet 2005;366(9495):1499-1503. Citado aquí

19. Cash R, Wikler D, Saxena A, et al. Casebook on ethical issues in international health research. Geneva: World Health Organization, 2009. Citado aquí

20. Katz R, Kornblet S, Arnold G, et al. Defining health diplomacy: changing demands in the era of globalization. The Milbnk Quarterly 2011;89(3):503-523. Citado aquí

CAPÍTULO 19 Elaboración de una propuesta para la financiación de la investigación

Steven R. Cummings
Deborah G. Grady
Stephen B. Hulley

El protocolo es el plan detallado del estudio por escrito. Redactar el protocolo obliga al investigador a organizar, aclarar y precisar todos los elementos del estudio, y esto fomenta el rigor científico y la eficacia del proyecto. Incluso si el investigador no precisa financiación para un estudio, sí es necesario un protocolo para guiar el trabajo y obtener la autorización ética del comité ético de la investigación clínica (CEIC). Una propuesta es un documento escrito con el fin de conseguir fondos de entidades patrocinadoras. Incluye descripciones de los objetivos del estudio, su significado, el abordaje de la investigación, las preocupaciones por los participantes humanos, el presupuesto, y otras informaciones administrativas y de apoyo que pida la agencia específica.

En este capítulo se describirá cómo elaborar una propuesta que consiga la financiación. Se centra en propuestas de investigaciones originales utilizando el formato propuesto por los National Institutes of Health (NIH), aunque las propuestas que se envían a la mayoría de las demás agencias financiadoras (como el Department of Veterans Affairs, los Centers for Disease Control, la Agency for Healthcare Research and Quality y fundaciones privadas) generalmente precisan un formato similar. Se pueden encontrar consejos excelentes sobre la redacción y la solicitud, la preparación de presupuestos y el envío de propuestas en la página web de los NIH (http://grants.nih.gov/grants/writing_application.htm).

▪ REDACCIÓN DE PROPUESTAS

La tarea de preparar una propuesta suele necesitar varios meses de organización, redacción y revisión. Los pasos siguientes pueden ayudar a que el proyecto arranque con un buen principio:

- Decida a dónde enviará la propuesta. Cada entidad financiadora tiene sus propias áreas de interés, procesos y requisitos para las propuestas. Por lo tanto, el investigador debe empezar por decidir a dónde enviará la propuesta, determinar el límite sobre el importe de la financiación, y obtendrá las directrices específicas sobre cómo preparar la propuesta y los plazos para esa entidad concreta. Los NIH son un buen lugar para comenzar, en http://grants.nih.gov/grants/oer.htm. Se pueden identificar las áreas de interés en las páginas web de los institutos individuales, en las que se describen sus prioridades. Se puede obtener información adicional sobre las áreas de interés actuales hablando con los administradores científicos de los institutos de los NIH, cuya información de contacto y áreas de responsabilidad se enumeran en la sección de anuncios de oportunidades de financiación *(Funding Opportunity Announcements)* de los NIH y en las páginas web de los distintos institutos.

- Organice un equipo y designe un director. La mayor parte de las propuestas las escribe un equipo de varias personas, que serán las que finalmente realizarán el estudio. Este equipo puede ser pequeño (el investigador y su mentor) o grande (colaboradores, un bioestadístico, un gestor administrativo, auxiliares de la investigación y personal de apoyo). Es importante que este equipo incluya o pueda acceder a la experiencia necesaria para diseñar el estudio.

  Uno de los miembros del equipo puede asumir la responsabilidad de dirigir el trabajo. Generalmente se trata del investigador principal (IP), que detentará la autoridad y la responsabilidad finales del estudio. El IP debe ejercer un liderazgo firme durante el desarrollo de la propuesta, delegando responsabilidades de relación y otras tareas, estableciendo fechas límite, realizando reuniones periódicas del equipo, garantizando que se realicen a tiempo todas las tareas necesarias, y asumiendo personalmente la calidad de la propuesta.

  El IP es muchas veces un científico con experiencia cuyo conocimiento y prudencia son útiles para las decisiones del diseño, y cuya trayectoria en estudios previos aumenta la probabilidad de éxito del estudio y, por lo tanto, de la financiación. Una vez dicho eso, los NIH animan a los nuevos investigadores a que soliciten financiación como IP, tienen oportunidades de financiación especial para ellos, y muchas veces dan preferencia a la financiación de sus propuestas (http://grants.nih.gov/grants/new_investigators/). La definición de los NIH de «nuevo investigador» es un científico que todavía no ha sido IP de una subvención de investigación de los NIH. Pero es muy probable que los nuevos IP reciban financiación si ya tienen cierta experiencia en la realización de investigaciones, bajo la guía de un científico con mayor experiencia y con financiación obtenida por ese compañero, con una subvención para el desarrollo profesional o con pequeñas subvenciones institucionales o de fundaciones. El historial de publicaciones, incluyendo haber sido primer autor en algunos de estos trabajos, es esencial para aportar datos de que el nuevo investigador tiene la posibilidad de ser un científico independiente con éxito, y está preparado y es capaz de dirigir la investigación.

  Un nuevo IP debe incluir en la solicitud de la subvención coinvestigadores que tengan un historial investigador de éxito en el área de interés para ofrecer guía sobre la realización del estudio y para mejorar la probabilidad de una revisión favorable. En ocasiones, esto se puede conseguir por el mecanismo de múltiples IP; los NIH permiten que haya más de un IP en las propuestas si los IP

aportan una experiencia diferente, pero complementaria, y si se definen con claridad sus diferentes funciones y responsabilidades (http://grants.nih.gov/grants/multi_pi/overview.htm).

- Siga las directrices de la entidad financiadora. Todas las entidades financiadoras proporcionan directrices por escrito que el investigador debe estudiar cuidadosamente antes de empezar a escribir la propuesta. Esta información contiene los tipos de investigación que se financiarán e instrucciones detalladas para organizar la propuesta, los límites de páginas, información sobre la cantidad de dinero que puede solicitarse y los elementos que deben incluirse en la propuesta.

  Sin embargo, estas directrices no contienen *toda* la información importante que el investigador debe conocer sobre el funcionamiento y las preferencias de las entidades financiadoras. En las fases tempranas de la elaboración de la propuesta es una buena idea discutir el plan con una persona de la agencia que pueda clarificar las preferencias de la agencia (como el ámbito y el detalle exigidos en la propuesta) y comentar si la agencia está interesada en el área de investigación prevista. Los NIH, otras agencias federales y las fundaciones privadas cuentan con administradores científicos («gestores de proyectos») cuyo trabajo consiste en ayudar a los investigadores a diseñar sus propuestas de forma que respondan mejor a las prioridades de financiación de la entidad. Puede ser muy útil contactar por correo electrónico o teléfono con el gestor de proyectos responsable de la correspondiente área de investigación para clarificar las directrices, los intereses y los procedimientos de revisión de la agencia. Posteriormente, una reunión con el gestor de proyectos en una conferencia científica que sea conveniente o al viajar cerca de la sede de la agencia es una buena forma de establecer una relación de trabajo que favorezca las propuestas financiables.

  Es útil hacer una lista de verificación de los detalles que se necesitan y revisar la lista repetidamente antes de enviar la propuesta. El rechazo de una propuesta por lo demás excelente por falta de cumplimiento de los detalles especificados es una experiencia frustrante y evitable. Los gestores de las subvenciones universitarias generalmente tienen listas de verificación que revisan antes de enviar una propuesta.

- Establezca un calendario y reúnase periódicamente. Un calendario para realizar las tareas de redacción mantiene una suave presión sobre los miembros del equipo, para que cumplan con sus obligaciones a tiempo. Además de abordar los componentes científicos especificados por la entidad financiadora, el calendario deberá tener en cuenta los requisitos administrativos de la institución donde se llevará a cabo la investigación. Las universidades a veces piden una revisión (que lleva su tiempo) del presupuesto y las subcontratas antes de que una propuesta pueda enviarse a una entidad financiadora, por lo que la fecha límite *real* para concluir una propuesta puede ser varios días o incluso varias semanas antes de la fecha límite establecida por la agencia. Dejar esos detalles para el final puede precipitar una crisis de última hora que afecte a una propuesta por lo demás bien realizada.

  Un calendario suele funcionar mejor si se especifican los plazos para los productos redactados, y si cada persona participa para finalizar sus propias tareas. Debe revisarse el calendario en reuniones periódicas del equipo de redacción, para comprobar que las tareas estén dentro del programa y que los plazos sean realistas.

- Encuentre una propuesta modelo. Es muy útil pedir prestada una propuesta reciente que haya tenido éxito dirigida a la entidad de la que se espera conseguir los fondos. Las solicitudes con éxito ilustran de un modo concreto el formato y el contenido de una buena propuesta. El investigador puede encontrar inspiración para nuevas ideas a partir del modelo, y diseñar y escribir una propuesta que sea más clara, lógica y persuasiva. También es una buena idea conseguir ejemplos de críticas por escrito que la entidad haya enviado anteriormente a propuestas que tuvieron y que no tuvieron éxito. Esto ilustrará los puntos clave que son importantes para los científicos que revisarán la propuesta. Esos ejemplos a menudo se pueden obtener de compañeros o de la oficina de investigación patrocinada de la institución del investigador.

- Trabaje a partir de un resumen. Hay que empezar estableciendo la propuesta como un resumen (tabla 19-1), que constituye un punto de partida para escribir y es útil para organizar las tareas que deben realizarse. Si son varias las personas que van a trabajar en la solicitud, el resumen ayuda a asignar responsabilidades para escribir partes de la propuesta. Uno de los obstáculos más habituales para crear un resumen es la sensación de que se debe elaborar todo el plan de la investigación antes de empezar a escribir la primera frase. El investigador debe dejar de lado esta idea y permitir que sus pensamientos fluyan al papel, para crear el material en bruto para editar, mejorar y obtener consejos de sus compañeros.

- Revise y repase repetidamente. Escribir una propuesta es un proceso iterativo; suele haber muchas versiones, cada una de las cuales refleja nuevas ideas, consejos y datos adicionales. Desde el comienzo del proceso de redacción de la propuesta, compañeros familiarizados con el tema en estudio y la agencia financiadora deben revisar críticamente los borradores. Debe prestarse especial atención a la importancia y el carácter innovador de la propuesta, la validez del diseño y los métodos, y la claridad de la redacción. Es mejor tener una crítica aguda y detallada antes de enviar la propuesta que ver cómo se rechaza el proyecto por no haber previsto y corregido los posibles problemas. Cuando la propuesta está casi lista para su envío, el paso final será revisar cuidadosamente su contenido, su formato, el cumplimiento de las directrices de la entidad y la ausencia de errores

tipográficos. Una redacción descuidada implica un trabajo descuidado y un liderazgo incompetente, y distrae significativamente de ideas, por lo demás, buenas.

Volver al principio

▪ ELEMENTOS DE UNA PROPUESTA PARA UNA SUBVENCIÓN IMPORTANTE

En la tabla 19-1 se muestran los elementos de una propuesta para una subvención de investigación importante, como una subvención NIH R01. Las solicitudes para otros tipos de subvenciones y contratos de los NIH, y de otras instituciones financiadoras, pueden precisar menos información o un formato diferente, y el investigador debe prestar mucha atención a las directrices de la agencia a la que se enviará la propuesta.

El comienzo

El título debe ser descriptivo y conciso. Proporciona la primera impresión y un recuerdo del objetivo principal de la investigación y del diseño del estudio. Por ejemplo, este título: «Ensayo aleatorizado de ecografía de alta frecuencia guiada por RM y ecografía simulada para tratar los fibroides sintomáticos», resume brevemente la pregunta de la investigación y el diseño del estudio. Se deben evitar frases innecesarias y vacías, como «Un estudio para determinar la…».

El resumen del proyecto es un resumen conciso del protocolo, que debe empezar con la pregunta y los fundamentos de la investigación, después debe establecer el diseño y los métodos, y debe concluir con una exposición de la importancia de los posibles hallazgos del estudio. El resumen debe ser informativo para las personas que trabajen en el mismo campo o en campos relacionados, y debe ser comprensible para un lector no especialista pero con conocimientos científicos. La mayor parte de las entidades exigen que el resumen tenga un número limitado de palabras, por lo que es mejor usar términos eficaces y descriptivos. Se deben realizar suficientes revisiones del resumen para garantizar que sea de primera calidad. Será la única página que leerán algunos revisores, y un recuerdo conveniente de las especificaciones de una propuesta para todos los demás. Debe, por lo tanto, incorporar las principales características del estudio propuesto, y revelar sus posibilidades y sus posibles consecuencias de una forma persuasiva.

Título

Resumen del proyecto

Partes administrativas

Presupuesto y justificación del presupuesto

Informes de los investigadores

Estructura física y recursos

Finalidades específicas

Estrategia de la investigación

Importancia

Innovación

Abordaje

Visión general

Justificación (justificación de la investigación prevista y datos preliminares)

Participantes del estudio

Criterios de selección

Diseño del muestreo

Planes para reunir participantes

Planes para optimizar el cumplimiento y el seguimiento

Procedimientos del estudio (si procede)

Aleatorización

Enmascaramiento

Mediciones

Principales variables predictivas (intervención si es un ensayo clínico)

Posibles variables de confusión

Variables de respuesta

Aspectos estadísticos

Método de análisis estadístico

Hipótesis, tamaño de la muestra y potencia

Contenido y calendario de las visitas del estudio

Gestión de los datos y control de calidad

Calendario y gráfico organizativo

Limitaciones y abordajes alternativos

Participantes humanos

Bibliografía

Apéndices y acuerdos de colaboración

Partes administrativas

Casi todas las entidades exigen una sección administrativa que incluya un presupuesto y una descripción de la idoneidad del personal, los recursos de la institución del investigador y el acceso a equipo, espacio y experiencia.

La sección sobre el presupuesto suele organizarse según las directrices de la entidad financiadora. Los NIH, por ejemplo, tienen un formato que solicita un presupuesto detallado para los primeros 12 meses, y un resumen del presupuesto para el período completo del proyecto propuesto (generalmente, 2-5 años). El presupuesto detallado de los 12 meses incluye las siguientes categorías de gastos: personal (con nombres y cargos de las personas que participan en el proyecto, el porcentaje de tiempo que cada uno dedicará al proyecto, y la cuantía que supondrá el salario y los beneficios adicionales para cada persona), los costes de los consultores, el equipo, los suministros, los viajes, la atención sanitaria de los pacientes, las alteraciones y renovaciones, los costes de consorcios/contractuales, y otros gastos (costes de teléfono, correo, multiconferencias, fotocopias, ilustraciones, publicaciones, libros y contratos de mantenimiento gratuitos).

El presupuesto no debe dejarse hasta el último minuto. Muchos elementos necesitan tiempo (para conseguir buenas estimaciones del coste del espacio, el equipo y el personal). Las universidades generalmente emplean administradores competentes cuyo trabajo es ayudar a los investigadores a preparar presupuestos y las otras partes administrativas de una propuesta. El mejor método es notificar a este administrador, lo antes posible, el plan de enviar una propuesta, y programar reuniones frecuentes con él para revisar el progreso y hacer un calendario para elaborar las secciones administrativas. Un administrador puede empezar a trabajar en cuanto se formule el esbozo de la propuesta, recomendando los importes de los distintos puntos del presupuesto y ayudando a garantizar que el investigador no pase por alto gastos importantes. Las instituciones cuentan con normas que deben seguirse y plazos que hay que cumplir, y un administrador con experiencia puede ayudar al investigador a anticiparse a esa normas, dificultades y posibles retrasos. El administrador también puede ser muy útil para elaborar el borrador del texto de la sección de justificación del presupuesto y los recursos, así como para recopilar los apéndices y otros materiales de apoyo para la propuesta.

En una justificación del presupuesto debe explicarse con detalle la necesidad de los importes solicitados para cada aspecto del presupuesto. Los salarios ocuparán generalmente la mayor parte del coste global de un proyecto de investigación clínica típico, por lo que es importante documentar la necesidad y las responsabilidades específicas de cada persona para justificar el esfuerzo porcentual necesario. Las descripciones completas, pero concisas, acerca de la labor de los investigadores y demás miembros del equipo de investigación no deben dejar duda a los revisores de la propuesta de que el esfuerzo calculado de cada integrante es esencial para que el proyecto tenga éxito.

Los revisores muchas veces están preocupados por los porcentajes de tiempo asignados a miembros esenciales del equipo de investigación. En ocasiones pueden criticarse las propuestas, porque personas fundamentales solo tienen una asignación de tiempo muy pequeña en el presupuesto y otros muchos compromisos distintos, lo que implica que pueden no ser capaces de dedicar la energía necesaria al estudio propuesto. Con más frecuencia, los revisores pueden oponerse a porcentajes que estén aumentados muy por encima de las necesidades que corresponden a la descripción del trabajo.

Incluso los presupuestos mejor planificados cambiarán a medida que lo hagan las necesidades del estudio, o se produzcan gastos o ahorros inesperados. En general, una vez que se haya concedido la subvención, se permite que el investigador gaste dinero de forma diferente a lo especificado en el presupuesto siempre que los cambios sean pequeños y los gastos se relacionen con los objetivos del estudio. Cuando el investigador desea desplazar dinero entre distintas partidas o realizar un gasto importante (mayor o menor) en la participación de los investigadores esenciales, puede necesitar la aprobación de la entidad financiadora. Esta generalmente aprueba las solicitudes razonables de actualizar el presupuesto siempre que el investigador no esté solicitando un aumento de los fondos totales.

Los NIH precisan una reseña biográfica breve de todos los investigadores y consultores que recibirán financiación gracias a la subvención. Las reseñas biográficas son resúmenes de cuatro páginas que siguen un formato especificado que incluye una declaración personal sobre en qué medida la experiencia del investigador hace que sea adecuado para la realización de este estudio, y enumera su formación académica y profesional, los cargos y empleos, los puestos honoríficos, un número limitado de publicaciones importantes, y las subvenciones y contratos de investigación importantes.

La sección de la propuesta sobre recursos disponibles para el proyecto puede incluir equipo técnico e informático, acceso a dispositivos especializados de técnicas de imagen o de medición, espacio para la consulta y el laboratorio, y recursos disponibles para facilitar la inclusión de los participantes, la recogida y la gestión de los datos y el almacenamiento de muestras. En la sección de recursos muchas veces se utilizan textos estándar: descripciones de propuestas previas o material suministrado por la institución, el centro o el laboratorio del investigador.

Objetivos específicos

Los objetivos específicos son descripciones de las preguntas de la investigación utilizando términos concretos para especificar la variable de respuesta deseada. Esta sección de una propuesta para los NIH debe ser concisa, porque está limitada a una única página. Y como esta es la página a la que prestan más atención muchos revisores, se debe escribir cuidadosamente y se debe revisar una y otra vez cuando se elabora la propuesta.

Un patrón habitual es comenzar con dos o tres párrafos cortos que resuman la información previa: la pregunta de la investigación y por qué es importante, los estudios que se han realizado y por qué no han resuelto el problema, y el abordaje previsto para responder a la pregunta en el estudio propuesto. A esto le sigue una declaración breve de los objetivos específicos, expresados como objetivos descriptivos tangibles y, siempre que sea posible, como hipótesis verificables.

Los objetivos se presentan en una secuencia lógica que el investigador adapta al estudio que quiere planificar. Puede comenzar con objetivos transversales para el período inicial, seguidos por objetivos relacionados con los hallazgos del seguimiento. Por otro lado, puede comenzar con objetivos que aborden mecanismos fisiopatológicos y acabar con objetivos que aborden variables de respuestas clínicas de salud pública. Un patrón que funciona especialmente bien para las subvenciones para el desarrollo profesional (denominado «investigación con métodos mixtos») comienza con objetivos cualitativos que pueden utilizar grupos de discusión para diseñar un instrumento o una intervención fundamental, a lo que siguen objetivos cuantitativos con variables predictivas, variables de respuesta y pruebas de hipótesis. Otro patrón adicional es comenzar con el objetivo más importante para resaltarlo; la secuencia de los objetivos muchas veces sirve como resumen para organizar secciones posteriores de la propuesta, por lo que este método tiene la ventaja de asignar al objetivo principal el primer lugar en todas las secciones de la propuesta, como el tamaño de la muestra y la potencia.

La sección de objetivos específicos muchas veces acaba con un breve párrafo final que resume de manera concisa las posibles consecuencias de los hallazgos del estudio sobre el conocimiento de la salud y la enfermedad, la práctica clínica, la salud pública o la investigación futura. El objetivo es hacer un alegato convincente que llevará a los miembros del comité de revisión que no eran revisores primarios o secundarios (y que pueden haber leído únicamente esta página de la propuesta) a respaldar una puntuación sobresaliente.

Estrategia de la investigación

El formato actual de los NIH limita la mayoría de los tipos de propuestas a 12 páginas para la presentación de la estrategia de la investigación, en tres secciones:

- La sección de la importancia, habitualmente dos o tres páginas, describe cómo los hallazgos del estudio harían avanzar el conocimiento científico, abordarían un problema importante o una barrera a los avances en esa área, mejorarían la práctica clínica o la salud pública, o influirían en las políticas sanitarias. Esta sección puede definir brevemente la magnitud del problema, resumir lo que ya se ha realizado, definir los problemas del conocimiento actual y mostrar en qué sentido el estudio propuesto generará avances en el campo.

- La sección de innovaciones, habitualmente una o dos páginas, señala aspectos en los que el estudio propuesto difiere de la investigación previa sobre el tema. Puede insistir en la posibilidad de documentar nuevos mecanismos de enfermedad, nuevos métodos de medición, poblaciones diferentes más extensas, nuevos métodos terapéuticos y preventivos, o nuevos abordajes del análisis de los datos. Las directrices de los NIH se centran en la forma en la que la investigación desviará los paradigmas actuales de investigación o práctica clínica mediante la utilización de conceptos, métodos o intervenciones innovadores. Una vez dicho eso, muchos estudios clínicos financiados dan lugar a mejorías y refinamientos tan solo incrementales de conceptos, métodos o intervenciones. Nuestro consejo es describir las nuevas características de la investigación *con exactitud*, sin afirmaciones grandilocuentes de que el estudio modificará los paradigmas o utilizará métodos totalmente innovadores.

- La sección del abordaje (antiguamente denominada «metodología») generalmente tiene de siete a nueve páginas de longitud. Ofrece los detalles del diseño y la realización del estudio, y es sometida a un intenso escrutinio por los revisores. Las directrices de los NIH sugieren que la sección de abordaje se debe organizar por objetivos específicos, y que debe incluir los componentes y la secuencia aproximada de la tabla 19-1 . Esta sección generalmente comienza con una breve visión de conjunto del abordaje, a veces acompañada por un diagrama esquemático o una tabla, para orientar al lector (tabla 19-2). La visión de conjunto debe describir con claridad el diseño del estudio y dar una breve descripción de los participantes en el estudio, las mediciones principales, cualquier intervención que se vaya a realizar, la duración del seguimiento y las variables de respuesta principales.

	VISITA DE SELECCIÓN	**ALEATORIZACIÓN EN LA VISITA**	**3 MESES**	**6 MESES**	**12 MESES**
Historia clínica	X	-	-	-	X
Presión arterial	X	X	X	X	X
Examen prostático	X	-	-	-	X
Antígeno prostático específico	X	-	-	-	X
Concentraciones de lípidos en sangre	-	X	X	X	X
Marcadores de inflamación	-	X	-	-	X
Densidad ósea	-	X	-	-	X
Marcadores de recambio óseo	-	X	X	-	X
Fuerza de prensión	-	X	X	X	X
Acontecimientos adversos	-	-	X	X	X

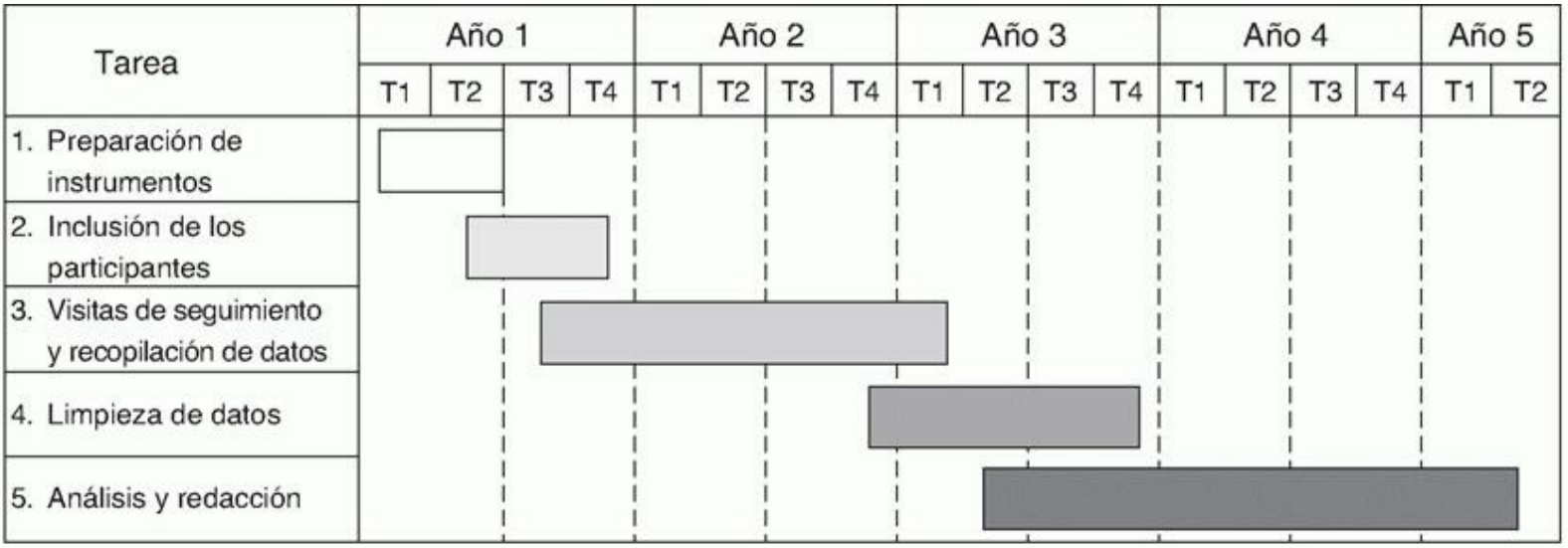
La sección de abordaje generalmente incluye una breve justificación de la investigación respaldada por datos preliminares, investigaciones previas realizadas por el investigador y su equipo que indican que el estudio propuesto tendrá éxito. Se debe insistir en la importancia del trabajo previo y los motivos por los que se debe continuar o ampliar. Los resultados de estudios piloto que confirman la importancia de la pregunta de la investigación y la viabilidad del estudio son importantes en muchos tipos de propuestas, especialmente cuando el equipo el investigador tiene poca experiencia previa con los métodos propuestos, cuando la pregunta es novedosa y cuando puede haber dudas sobre la viabilidad de los procedimientos propuestos o de la inclusión de los participantes. Esta es una oportunidad de demostrar que el investigador y su equipo tienen la experiencia y los conocimientos específicos necesarios para realizar el estudio.

Otros componentes específicos de la sección de abordaje ya se han analizado. La sección de participantes del estudio (cap. 3) debe definirse con claridad y ofrecer la justificación de los criterios de inclusión y exclusión, y debe especificar el método de muestreo. Es importante describir cómo se reunirán los participantes del estudio y garantizar a los revisores que los investigadores son capaces de incluir el número deseado de participantes en el estudio. Se deben ofrecer planes para optimizar el cumplimiento de la intervención del estudio (si procede) y de las visitas del estudio.

La sección de abordaje debe incluir una descripción de los procedimientos del estudio importantes, como la aleatorización y el

enmascaramiento. La sección de mediciones del estudio (cap. 4) debe describir con claridad cómo se medirán las variables predictivas y de respuesta y posibles variables de confusión, y en qué momento del estudio se realizarán estas mediciones, además de cómo se aplicarán las intervenciones, y como se evaluará y medirá la variable de respuesta principal.

La sección de estadística habitualmente comienza con los planes de análisis, organizados por objetivo específico. El plan se puede presentar en la secuencia lógica; por ejemplo, primero las tabulaciones descriptivas y después el abordaje para analizar las asociaciones entre las variables. A esto le sigue un análisis del tamaño de la muestra y la potencia (caps. 5 y 6), que debe comenzar con una declaración de la hipótesis nula para el objetivo, y que determinará el tamaño de la muestra del estudio. Las estimaciones del tamaño de la muestra y la potencia se basan en suposiciones sobre la magnitud de las asociaciones que probablemente se detecten y la precisión de las mediciones que se realizarán. Estas suposiciones deben estar justificadas con referencias de la bibliografía publicada o en trabajos preliminares que respalden estos juicios. Muchas veces es útil incluir una tabla o una figura que muestre cómo influyen en el tamaño de la muestra las variaciones de la magnitud del efecto, la potencia u otros supuestos, para demostrar que el investigador ha realizado elecciones razonables. La mayoría de los paneles de revisión de los NIH asigna mucha importancia a la sección de estadística, por lo que es una buena idea que un estadístico participe en la redacción de este componente.



▪FIGURA 19-1. Calendario hipotético.
Herramientas de imágenes

Es útil incluir una tabla que presente las visitas del estudio o los contactos con los participantes, el momento de realización de las visitas, y qué procedimientos y mediciones se realizarán en cada visita. Esta tabla ofrece una visión de conjunto breve de todas las actividades del estudio (v. tabla 19-2). Las descripciones de la gestión de los datos y del control de calidad (caps. 16 y 17) deben abordar cómo se recogerán, almacenarán y editarán los datos del estudio, junto con los planes para maximizar la calidad de los datos y la seguridad.

La propuesta debe ofrecer un plan de trabajo y un calendario realistas en los que se incluyan las fechas en las que comenzará y finalizará cada una de las fases principales del estudio (fig. 19-1). Se pueden preparar calendarios similares para los patrones de dotación de personal y otros componentes del proyecto. Para estudios extensos se puede elaborar un gráfico organizativo en el que se describa el equipo de la investigación y en el que se pueden indicar los niveles de autoridad y responsabilidad y las líneas de notificación, y se mostrará cómo funcionará el equipo.

Aunque no es una sección necesaria, puede ser útil incluir una discusión de las limitaciones de la investigación propuesta y los abordajes alternativos. En lugar de ignorar los posibles problemas, el investigador debe decidir abordarlos de manera explícita, y discutir las ventajas y las desventajas de los posibles compromisos para llegar al plan elegido. Si se señalan los retos importantes y las posibles soluciones, las posibles críticas a la solicitud se pueden transformar en puntos fuertes. Sin embargo, es un error insistir en exceso en estos problemas, porque esto puede llevar a un revisor a centrarse de manera excesiva en los aspectos más débiles de la propuesta. El objetivo es tranquilizar al revisor de que el investigador ha previsto todos los posibles problemas importantes y tiene un abordaje realista y meditado para afrontarlos.

Componentes finales de una propuesta importante

La sección sobre participantes humanos aborda los temas éticos que surgen del estudio, tratando los aspectos de seguridad, privacidad y confidencialidad. Esta sección indica planes específicos para informar a los posibles participantes de los riesgos y beneficios y para obtener su consentimiento para participar (cap. 14). También describe la inclusión de mujeres, niños y participantes de grupos minoritarios, tal y como se exige para las propuestas al NIH, y debe justificar la exclusión de cualquiera de estos grupos.

La bibliografía envía un mensaje sobre la familiaridad del investigador con un determinado campo. Debe ser completa, pero parca y actualizada, no una lista exhaustiva y no seleccionada. Debe citarse cada referencia exactamente; los errores en estas citas o la interpretación errónea de los trabajos serán contemplados negativamente por los revisores que estén familiarizados con el campo de esa investigación.

Para algunos tipos de propuestas, los apéndices pueden ser útiles para ofrecer la información técnica detallada y el material de apoyo mencionados brevemente en el texto. (Sin embargo, para evitar que el uso de apéndices permita saltarse los límites del número de páginas de las propuestas, los NIH limitan estrictamente su uso.) En los apéndices se pueden incluir instrumentos para la recogida de datos (como cuestionarios) y protocolos clínicos, y hasta tres manuscritos y resúmenes que hayan sido aceptados, pero que todavía no se hayan publicado. Los revisores principales y secundarios son los únicos miembros del comité de revisión que reciben los apéndices. Por lo tanto, en la propuesta principal se debe resumir brevemente todo aquello que sea importante.

Debe describirse el uso propuesto y el valor de cada consultor, acompañado de una carta de acuerdo firmada por dicho consultor y una copia de su currículo. (Los investigadores que recibirán financiación en forma de salario con la subvención no deben enviar cartas.) También se deben incluir otras cartas de apoyo, como las que hayan elaborado personas que ofrecerán el acceso al equipo o los recursos. Debe incluirse una explicación de los acuerdos administrativos y programáticos entre la organización solicitante y las instituciones colaboradoras y los laboratorios, acompañada por cartas de compromiso de los funcionarios responsables dirigidas al investigador.

Volver al principio

▪ CARACTERÍSTICAS DE LAS BUENAS PROPUESTAS

Una buena propuesta para la financiación de una investigación consta de varios atributos. En primer lugar está la calidad científica de la estrategia de la investigación: debe basarse en una buena pregunta de la investigación, usar un diseño y unos métodos rigurosos y viables, y disponer de un equipo investigador con experiencia, habilidades y compromiso para llevarlo a cabo. El segundo aspecto es la claridad de la presentación; una propuesta concisa y atractiva, bien organizada, redactada cuidadosamente, presentada de forma atractiva y sin errores hace que el lector crea que la realización de la investigación tendrá una calidad similarmente elevada.

Los revisores muchas veces se ven superados por un gran número de propuestas, de modo que los méritos del proyecto deben presentarse de un modo que no pueda pasar por alto ni siquiera tras una lectura rápida y superficial. Un resumen claro que siga los objetivos específicos, las secciones cortas con subapartados significativos, y el uso de tablas y figuras para interrumpir párrafos largos, pueden dirigir el conocimiento por el revisor de las características más importantes de la propuesta. Las directrices actuales de los NIH recomiendan comenzar los párrafos con una frase sobre el tema en negrita que constituya el punto fundamental, lo que permitirá que los revisores agobiados comprendan los elementos principales de la propuesta realizando un rastreo rápido de las frases sobre el tema. Es importante tener en consideración los diversos puntos de vista y la distinta experiencia de los revisores, incluyendo suficientes detalles para convencer a un revisor experto de la importancia y la sofisticación del trabajo propuesto, a la vez que se sigue atrayendo al mayor número de revisores que no estén familiarizados con el área de la investigación.

La mayoría de los revisores rechazan la sobrevaloración y otras formas torpes de búsqueda de subvenciones. Las propuestas que exageran la importancia del proyecto y dan un valor excesivo a lo que pueden lograr generarán escepticismo. Escribir con entusiasmo es bueno, pero el investigador debe ser realista sobre las limitaciones del proyecto. Los revisores son expertos en identificar posibles problemas en el diseño o la viabilidad.

Una última revisión científica por científicos expertos que no han participado en el desarrollo de la propuesta, en un momento en que todavía pueden realizarse cambios importantes, puede ser extraordinariamente útil para la propuesta, así como una experiencia colectiva provechosa. También es útil que haya alguien con excelentes habilidades de redacción que complemente con consejos sobre el estilo y la claridad a los programas de revisión ortográfica y gramatical de los procesadores de textos.

Volver al principio

▪ BÚSQUEDA DE APOYO PARA LA INVESTIGACIÓN

Los investigadores deben estar atentos a las oportunidades para realizar buenas investigaciones sin propuestas formales de financiación. Por ejemplo, un investigador novel puede analizar personalmente series de datos que otros han recopilado, o disponer de pequeñas cantidades de tiempo de científicos con experiencia o de su departamento para realizar estudios pequeños. Realizar una investigación sin la financiación de propuestas formales es más rápido y sencillo, aunque tiene el inconveniente de que el proyecto necesariamente tiene

un ámbito limitado. Además, las instituciones académicas a menudo toman decisiones sobre las mejoras de la carrera profesional en parte según el antecedente de consecución de fondos externos para la investigación de un científico.

He aquí cuatro categorías principales de fondos para investigación médica:

- Gubernamentales (NIH, Department of Veterans Affairs, Centers for Disease Control and Prevention [CDC], Agency for Healthcare Research and Quality [AHRQ], Patient Centered Outcomes Research Institute [PCORI], Ministerio de Defensa [Department of Defense, DOD], y otras muchas agencias federales, estatales y municipales).

- Fundaciones, sociedades profesionales, como la American Heart Association (AHA) y la American Cancer Society (ACS), y donantes individuales.

- Corporaciones (fundamentalmente, compañías farmacéuticas y fabricantes de dispositivos); y

- Recursos internos (p. ej., la universidad del investigador).

  Lograr el apoyo de estas fuentes es un proceso complejo y competitivo, que favorece a los investigadores con experiencia y tenacidad, y los investigadores noveles deben buscar un mentor con estas características. En las secciones siguientes se analizarán varias de las fuentes de financiación más importantes.

Subvenciones y contratos de los NIH

Los NIH ofrecen muchos tipos de subvenciones y contratos. Las subvenciones «R» (R01, y las de menor cuantía, R03 y R21) apoyan proyectos de investigación concebidos por el investigador sobre un tema de su elección, o escritos en respuesta a una solicitud de uno de los NIH (véase www.nimh.nih.gov/research-funding/grants/research-grants-r.shtml). Las subvenciones «K» (K23, K01, K08, K24, y las K12 y KL2, concedidas a nivel local) son un excelente recurso que ofrece financiación en forma de salario para la transformación y el desarrollo de la carrera profesional de investigadores noveles, además de una financiación modesta para la investigación (véase www.grants.nih.gov/training/careerdevelopmentawards.htm/).

Las propuestas a iniciativa del instituto se diseñan para estimular la investigación en áreas designadas por comités de los NIH, y adoptan la forma de peticiones de propuesta (*Requests for Proposals*, RFP) y peticiones de solicitud (*Requests for Applications*, RFA). En respuesta a una RFP, el investigador se compromete por contrato a realizar unas actividades de investigación específicas determinadas por los NIH. Bajo una RFA, el investigador dirige una investigación en un área temática definida por los NIH, con una pregunta de la investigación específica y un plan de estudio que ha propuesto él mismo. Las RFP usan un mecanismo de contrato para pagar a la institución del investigador los costes que supone la consecución de los objetivos planificados, y las RFA usan un mecanismo de subvención para apoyar actividades más abiertas.

Después de enviar una propuesta, la solicitud pasa por un proceso de revisión, que incluye una revisión administrativa inicial por el personal de los NIH, una revisión externa por un grupo de científicos, recomendaciones sobre la financiación por el comité asesor del instituto, y la decisión final sobre la financiación por el director del instituto. Las solicitudes de subvención suele revisarlas una de las muchas «secciones de estudio» de los NIH, grupos de revisores científicos con un área específica de experiencia en investigación y procedentes de instituciones investigadoras de todo el país. En la página web de los NIH se presenta una lista de secciones de estudio y de sus miembros actuales.

El proceso de los NIH para la revisión y las propuestas de financiación se describe en cms.csr.nih.gov. Cuando el investigador envía una solicitud de financiación, es asignado por el Center for Scientific Review (CSR) del NIH a una sección de estudio particular (fig. 19-2). Las propuestas se asignan a un revisor principal y a dos o más revisores secundarios, cada uno de los cuales ofrece una puntuación separada desde 1 a 9 para la importancia, la innovación, la metodología, los investigadores y el entorno, y después una puntuación total de las probables consecuencias del estudio. Una puntuación de «1» indica una solicitud excepcionalmente fuerte que prácticamente no tiene ningún punto débil, y un «9» es una solicitud con aspectos débiles graves e importantes y muy pocos puntos fuertes. Las puntuaciones de los revisores designados se comunican a la sección de estudio, y las propuestas con puntuaciones en la mitad superior son discutidas con todo el comité; el resto se «separan» (no se discuten), y algunas se aplazan hasta el siguiente ciclo 4 meses después, pendientes de la clarificación de aspectos que no estaban claros. Después de la discusión, los revisores asignados proponen de nuevo puntuaciones (las puntuaciones pueden haber cambiado como consecuencia de la discusión), y después todos los miembros del comité asignan puntuaciones mediante votación secreta. Estas puntuaciones se promedian, se multiplican por 10 para obtener una puntuación total que va desde 10 (mejor) hasta 90 (peor), y se utilizan en cada uno de los institutos para priorizar las decisiones de financiación.



▪FIGURA 19-2. Visión general de las fuentes de financiación de los NIH y las fundaciones, y de sus mecanismos.
Herramientas de imágenes

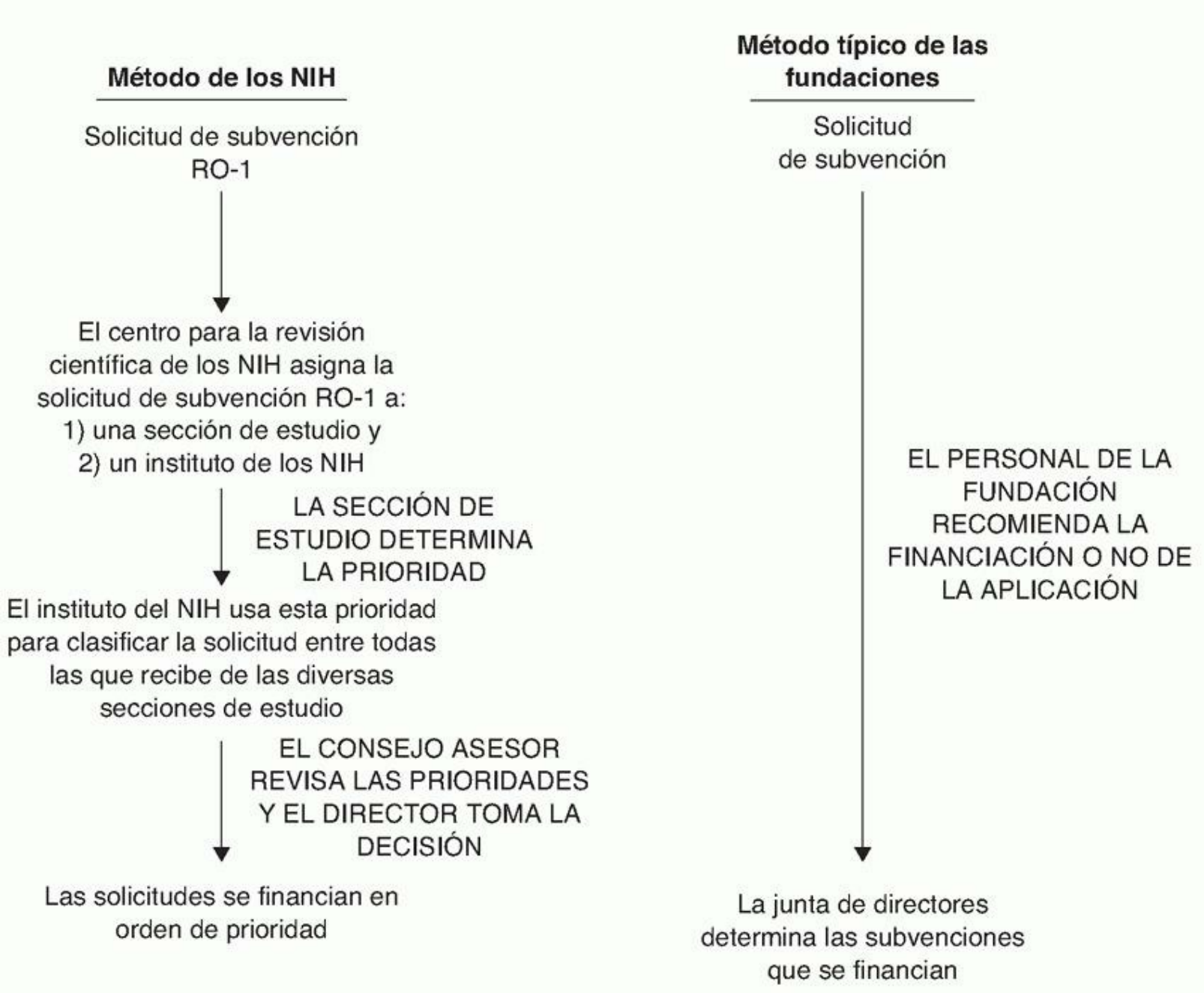
El investigador debe decidir por adelantado, con el consejo de compañeros con mayor experiencia, qué sección de estudio sería la mejor opción para revisar la propuesta. Las secciones de estudio varían mucho no solo en el área de estudio, sino también en la experiencia de los revisores y la calidad de las solicitudes que compiten. Aunque no se puede controlar por completo la asignación a una sesión de estudio, el investigador puede influir en la asignación, implicando al gestor de proyectos con el que ha trabajado cuando le envió la solicitud.

Además de asignar cada una de las solicitudes de financiación a una sección de estudio particular, el CSR también la asigna a un instituto (o centro) particular de los NIH. Posteriormente, cada instituto financia las subvenciones asignadas al mismo, en orden de puntuación de prioridad, matizada por la revisión por un consejo asesor, y en ocasiones invalidada por el instituto (fig. 19-3). Las propuestas de nuevos investigadores que todavía no han recibido financiación para la investigación de los NIH se financian con puntuaciones y valores de corte de percentiles algo más benévolos que las de investigadores ya consolidados. Si una solicitud tiene interés para más de un instituto, en ocasiones los institutos acuerdan compartir la financiación.

Después de la revisión de una solicitud, el investigador recibe una notificación por escrito de la decisión de la sección de estudio. Esta declaración resumen incluye la puntuación y comentarios y críticas detallados de los miembros del comité que revisaron la solicitud.

Las solicitudes al NIH que no son financiadas, como ocurre a menudo la primera vez, se pueden revisar y volver a enviar tan solo una vez. Si las críticas y puntuaciones de los revisores sugieren que se puede conseguir que la solicitud sea más atractiva para el comité, entonces una versión revisada puede tener una probabilidad excelente de obtener financiación cuando se vuelva a enviar. (Puede ser más difícil entusiasmar a los revisores si indican que la propuesta carece de innovación o importancia.) Los gestores de proyectos del correspondiente instituto habitualmente asisten a las reuniones de la sección de estudio, y es importante discutir la revisión con uno de ellos poco después de la reunión, porque habitualmente ya se ha elaborado el borrador de los comentarios escritos antes de la reunión, y

puede no reflejar los problemas que plantearon los miembros de la sección de estudio y que llevaron a las revisiones de las puntuaciones.



▪FIGURA 19-3. Procedimientos de los NIH y las fundaciones para revisar las solicitudes de subvención.
Herramientas de imágenes

No es necesario que un investigador haga automáticamente todos los cambios propuestos por los revisores, aunque debe adoptar las revisiones que satisfagan las críticas de los revisores siempre que sea posible, y debe justificar cualquier decisión de no hacerlo. Los NIH limitan las respuestas a los revisores a una introducción en una única página en la que se describan los cambios que se han hecho a la propuesta revisada. Un buen formato para la introducción es resumir con brevedad cada una de las críticas principales de la declaración del resumen en negrita o cursiva, y abordarla con una declaración breve del cambio consiguiente de la propuesta. Para ayudar a los revisores a centrarse en estas revisiones, los cambios se deben marcar, por ejemplo, por una línea vertical en el margen izquierdo del texto.

Subvenciones de fundaciones y sociedades profesionales

Las fundaciones privadas (como The Robert Wood Johnson Foundation) suelen limitar su financiación a determinadas áreas de interés. Algunas fundaciones sobre alguna enfermedad y las sociedades profesionales (como la American Heart Association y la American Cancer Society) también patrocinan programas de investigación, muchos de los cuales están diseñados para apoyar a investigadores noveles. La cantidad total de la subvención a la investigación es mucho menor que la que proporcionan los NIH, y la mayor parte de las fundaciones tienen el propósito de usar este dinero para financiar proyectos meritorios que tengan temas o métodos que es poco probable que sean financiados por los NIH. Algunas fundaciones ofrecen galardones para el desarrollo de la carrera profesional que están

centrados en áreas específicas, como la calidad de la asistencia sanitaria. El Foundation Center (http://fdncenter.org/) mantiene un directorio de fundaciones e información de contacto en el que se pueden realizar búsquedas, junto con los consejos sobre cómo redactar propuestas eficaces a las fundaciones. Las decisiones sobre la financiación siguen unos procedimientos que varían según la institución, pero suelen responder rápidamente a propuestas relativamente cortas (v. fig. 19-3). Las decisiones se toman, con frecuencia, mediante un proceso ejecutivo, en lugar de una revisión experta; habitualmente, el personal de la fundación realiza una recomendación, que es ratificada por una junta de directores.

Para determinar si una fundación podría estar interesada en una propuesta concreta, un investigador debe consultar con su mentor y comprobar la página web de la fundación. En esta se describirán, generalmente, los objetivos y propósitos de la fundación, y a menudo se enumerarán los proyectos que ha patrocinado recientemente. Si parece que la fundación podría ser una fuente de apoyo adecuada, lo mejor es contactar con el correspondiente miembro de aquella para describirle el proyecto, determinar su posible interés y lograr una orientación sobre cómo enviar la propuesta. Muchas fundaciones quieren que los investigadores envíen una carta que describa los antecedentes y los principales objetivos del proyecto, los méritos de los investigadores, y la duración y el coste aproximados de la investigación. Si la carta es lo suficientemente interesante, la fundación puede solicitar una propuesta más detallada.

Apoyo de la industria a la investigación

Las corporaciones que fabrican fármacos y dispositivos son una importante fuente de financiación, especialmente para ensayos aleatorizados de nuevos tratamientos. Las grandes compañías suelen aceptar solicitudes para investigaciones iniciadas por el investigador que pueden incluir pequeños estudios sobre los efectos o los mecanismos de acción de un tratamiento, o estudios epidemiológicos sobre afecciones de interés para la compañía. Con frecuencia suministran el fármaco y un placebo para un ensayo clínico propuesto por un investigador que tiene interés para la compañía. También pueden ofrecer pequeñas subvenciones para apoyar programas educativos en áreas de su interés. Sin embargo, con mucho, la mayor forma de apoyo de la industria a la investigación clínica es mediante contratos con los investigadores principales de centros clínicos para la inclusión de participantes en ensayos multicéntricos en los que se evalúen nuevos fármacos o dispositivos. Estos extensos estudios a veces son diseñados y gestionados por un centro coordinador académico, aunque habitualmente son dirigidos por el promotor corporativo, muchas veces mediante un contrato con una organización de investigación clínica (OIC).

Las solicitudes de apoyo a la investigación o a programas formativos, o para participar como centro de investigación en un ensayo clínico, suelen empezar entrando en contacto con el representante regional de la compañía. Si esta está interesada en el tema, puede que se le pida al investigador que envíe una solicitud relativamente breve y que cumplimente un formulario de presupuesto y otros formularios. A menudo, las compañías dan preferencia a las solicitudes de «líderes de opinión», médicos o investigadores conocidos que han participado en investigación o son asesores de la compañía, cuyas opiniones pueden influir en el modo en que otros médicos recetan fármacos o usan dispositivos. Por lo tanto, un investigador joven que busque el apoyo de la industria deberá, generalmente, lograr la ayuda de un mentor conocido para contactar con la compañía y enviar la solicitud.

Los contratos para incluir participantes en ensayos clínicos generalmente suponen el pago a los investigadores principales de los centros de investigación de una tarifa fija por cada participante incluido en un ensayo multicéntrico, y se cierra la inclusión en el ensayo cuando se ha cumplido el objetivo deseado en todo el estudio. Un investigador puede incluir suficientes participantes como para recibir fondos que superen sus costes, en cuyo caso puede quedarse con el sobrante, pero también puede perder dinero si incluye un número demasiado bajo de participantes para poder pagar al personal y los gastos generales del ensayo. Antes de decidir si va a participar como centro en estos ensayos multicéntricos, el investigador debe asegurarse de que el contrato pueda ser autorizado por las oficinas administrativas y el comité ético de la investigación clínica de su institución a tiempo para incluir suficientes participantes antes de que finalice la inclusión.

La financiación por la industria, particularmente por los departamentos de *marketing,* a menudo se canaliza hacia temas y actividades que pretenden aumentar las ventas del producto de la compañía. Los hallazgos de los ensayos gestionados por las compañías farmacéuticas generalmente son analizados por estadísticos de la compañía, y, en ocasiones, sus redactores médicos son quienes elaboran los borradores. Generalmente se selecciona a varios investigadores principales de distintos centros para que sean los coautores en publicaciones con revisión externa. Las regulaciones federales exigen que los autores tengan acceso a los datos (incluyendo el derecho a que realicen estudios de los datos de todo el estudio), que hagan contribuciones importantes a los manuscritos y que asuman la responsabilidad de las conclusiones; animamos a los investigadores principales de los distintos centros a que busquen roles de autoría para ellos mismos y para sus colaboradores y, si tienen éxito, para que cumplan estos requisitos de autoría. En la situación ideal, los planes de análisis, los manuscritos y las presentaciones de los estudios multicéntricos deben ser revisados y autorizados por un comité de publicaciones que tenga directrices elaboradas y que esté formado por una mayoría de miembros que no sean empleados del promotor.

Una ventaja del apoyo corporativo es que es el único modo práctico de encauzar algunas preguntas de investigación. No habría otra fuente de financiación, por ejemplo, para estudiar un nuevo antibiótico que no se encuentra todavía en el mercado. Otra ventaja es la relativa rapidez con la que se puede conseguir esta fuente de financiación; las decisiones sobre pequeñas propuestas por parte de los investigadores se toman en unos meses, y las compañías farmacéuticas están ansiosas por captar investigadores cualificados para

participar en sus ensayos clínicos multicéntricos. Los científicos de la compañía generalmente tienen mucha experiencia en el tratamiento y la metodología de la investigación, lo que puede ser útil para la planificación de los análisis y la interpretación de los resultados. Además, la mayor parte de las compañías farmacéuticas recompensan por mantener una reputación de integridad que fomenta sus relaciones con la vigilante FDA en Estados Unidos y su lugar de cara al público. La experiencia en investigación, el apoyo estadístico y los recursos económicos que proporcionan pueden mejorar la calidad de la investigación.

Apoyo interno

Las universidades cuentan, a menudo, con fondos para la investigación local para sus propios investigadores. Las subvenciones con estos fondos suelen limitarse a cantidades relativamente pequeñas, pero suele poder disponerse de ellas con mayor rapidez (semanas o meses), y para una mayor proporción de solicitantes, que las subvenciones de los NIH o de fundaciones privadas. Los fondos internos pueden estar restringidos a determinados fines, como estudios piloto, que pueden conducir a lograr una financiación externa, o adquisición de equipo. Estos fondos a veces están destinados a jóvenes del cuerpo docente y ofrecen una oportunidad única para que un nuevo investigador adquiera la experiencia de dirigir un proyecto financiado.

Volver al principio

▪ RESUMEN

- Una propuesta es una versión ampliada del plan de un estudio por escrito y detallado (protocolo) que se utiliza para solicitar financiación, y también contiene la información presupuestaria, administrativa y auxiliar que exige la agencia financiadora.
- Un investigador que está trabajando en una propuesta de investigación debe empezar obteniendo el consejo de compañeros con experiencia para la pregunta de la investigación que intentará resolver y la elección de la entidad financiadora. Los siguientes pasos serán estudiar las directrices de esa entidad y contactar con un administrador científico de esta para lograr asesoramiento.
- El proceso de redactar una propuesta, que a menudo tarda más de lo esperado, conlleva organizar un equipo con la experiencia necesaria; designar un investigador principal (IP); resumir la propuesta para que se adapte estrictamente a las directrices de la entidad, establecer un calendario para productos escritos; buscar un modelo de propuesta; y revisar el progreso en reuniones regulares. La propuesta deben revisarla compañeros apropiados, y debe repasarse con frecuencia y «pulirse» al final, prestando atención a los detalles.
- Los principales elementos de una propuesta son el resumen, las partes administrativas centradas alrededor del presupuesto, la justificación del presupuesto, breves reseñas biográficas y recursos, los objetivos específicos, que son muy importantes, y la estrategia de la investigación, con sus secciones de importancia, innovaciones y abordaje, incluyendo la investigación previa por el investigador.
- Una buena propuesta no solo necesita una buena pregunta de la investigación, un plan del estudio y un equipo de investigación, sino también una presentación clara: la propuesta debe comunicarse de forma clara y concisa, siguiendo un esquema lógico e indicando las ventajas y los inconvenientes de las concesiones en el plan de estudio. Los méritos de la propuesta deben resaltarse utilizando subencabezamientos, tablas y diagramas, para que no los pase por alto un revisor ocupado.
- Cuatro son las principales fuentes que sostienen la investigación clínica:
    - Los NIH y otros organismos gubernamentales son los principales proveedores de apoyo, mediante un sistema complejo de revisión especializada y administrativa, que se mueve lentamente, aunque aporta una amplia variedad de subvenciones y contratos de investigación y para el desarrollo de la carrera profesional.
    - Las fundaciones y sociedades están muchas veces interesadas en prometedoras investigaciones que escapan de la financiación de los NIH, y tienen procedimientos de revisión más rápidos, pero más localistas que los NIH.
    - Los fabricantes de fármacos y dispositivos constituyen una gran fuente de financiación, que, en su mayor parte, está canalizada hacia estudios de la propia compañía sobre nuevos fármacos y dispositivos médicos; sin embargo, las compañías valoran la relación con científicos importantes y apoyan algunas investigaciones iniciadas por los investigadores.
    - Los fondos internos de la institución del investigador tienden a contar con tasas de financiación favorables para lograr pequeñas cantidades de dinero de forma rápida, y son un excelente primer paso para estudios piloto y nuevos investigadores.

Volver al principio

Ejercicios
NA

Capítulo 1 Cómo empezar: anatomía y fisiología de las investigaciones clínicas

- En el apéndice 1 se presenta un resumen del estudio *Early Limited Formula* (ELF) realizado en dos centros médicos académicos de California con el objetivo de fomentar la lactancia materna en recién nacidos que habían perdido ≥ 5 % del peso corporal. En este estudio aleatorizado, la proporción de madres que comunicaron a un entrevistador que desconocía la asignación al tratamiento que daban exclusivamente lactancia materna a los 3 meses era del 79 % en el grupo de fórmula limitada temprana, en comparación con el 42 % en el grupo testigo ($p = 0,02$) (Flaherman y cols., *Pediatrics* 2013;131 [in press]). Para cada una de las siguientes afirmaciones, indique: *1)* si es una inferencia de validez interna o de validez externa; *2)* si usted piensa que es una inferencia válida, y *3)* los motivos por los que podría *no* ser válida.
    - Para las mujeres de este estudio, la provisión de una fórmula limitada temprana aumentó las tasas de lactancia materna a los 3 meses.
    - La provisión de una fórmula limitada temprana a lactantes con una pérdida de peso ≥ 5 % en las primeras 36 h nacidos en un hospital comunitario de Boston probablemente llevará a mayores tasas de lactancia materna a los 6 meses de edad.
    - De acuerdo con los resultados de este estudio, es probable que un proyecto internacional para suministrar esta fórmula a la mayoría de los recién nacidos mejore el éxito de la lactancia materna y la salud de los recién nacidos y de sus madres.
- Para cada uno de los siguientes resúmenes extraídos de estudios publicados, *escriba una única frase* que especifique el diseño y la pregunta de la investigación, y que incluya las principales variables predictivas y de respuesta y la población en la que se realizó el muestreo.
    - Investigadores de Winston-Salem, Carolina del Norte, realizaron una encuesta a una muestra aleatoria de 2 228 estudiantes del instituto de educación secundaria local sobre la frecuencia con la que habían visto combates de lucha en la televisión en las 2 semanas previas, y 6 meses después preguntaron a los mismos estudiantes por las peleas en la escuela y en las citas. Las posibilidades ajustadas de referir peleas en una cita aumentaron un 14 % con cada episodio de lucha que referían los estudiantes que habían visto 6 meses antes (DuRant y cols., *Pediatrics* 2006;118:e265-272).
    - Para evaluar si la magnitud de la lactancia materna protege a las mujeres frente al cáncer de ovario, los investigadores estudiaron a 493 mujeres chinas con cáncer de ovario recién diagnosticado y otras 472 mujeres hospitalizadas, todas las cuales habían dado lactancia materna al menos a un hijo. Encontraron una relación de dosis-respuesta entre los meses totales de lactancia materna y la reducción del riesgo de cáncer de ovario. Por ejemplo, las mujeres que habían dado de mamar durante al menos 31 meses tenían una razón de posibilidades de 0,09 (IC del 95 %: 0,04-0,19) en comparación con las mujeres que habían dado de mamar menos de 10 meses (Su y cols., *Am J Clin Nutr* 2013;97:354-359).
    - Para ver si la asociación entre la ingesta de grasas saturadas de la dieta y la reducción de la concentración de espermatozoides en hombres estériles se extendía a la población general, investigadores daneses recogieron muestras de semen y cuestionarios de frecuencia de alimentos de hombres jóvenes que dieron su consentimiento en el momento de su evaluación para el servicio militar. Encontraron una relación dosis-respuesta significativa entre la ingesta de grasas saturadas con la dieta referida por los participantes y la reducción de la concentración de espermatozoides (es decir, una concentración de espermatozoides un 41 % menor [IC del 95 %: 4 %-64 %] en el cuartil superior de ingesta de grasa saturada en comparación con el inferior) (Jensen y cols., *Am J Clin Nutr* 2013;97:411-418).
    - No hay ningún tratamiento farmacológico eficaz conocido para el ~20 % de pacientes con diarrea por *Clostridium difficile* que recidiva después del tratamiento con antibióticos. Investigadores de Ámsterdam estudiaron a pacientes ≥ 18 años de edad que tuvieron una recurrencia de la diarrea por *C. difficile* después de al menos un ciclo de tratamiento antibiótico adecuado. Se les asignó aleatoriamente (sin enmascaramiento) a uno de tres regímenes: un ciclo de 5 días de vancomicina seguido por lavado intestinal e infusión de una suspensión de heces de donante voluntario a través de una sonda gastroduodenal, o un ciclo estándar de 14 días de vancomicina con o sin lavado intestinal el día 4 o 5. El estudio se suspendió precozmente después de que un análisis preliminar mostrara que la tasa de curación sin recurrencia durante 10 semanas era de 13 de 16 (81 %) en el grupo de heces de donante, en comparación con 4 de 13 con vancomicina sola y 3 de 13 con vancomicina más lavado ($p < 0,001$ para ambas comparaciones) (van Nood y cols., *N Engl J Med* 2013;368:407-415).

Capítulo 2 Elaboración de la pregunta de la investigación y desarrollo del plan del estudio

- Considere la pregunta de la investigación: «¿Cuál es la relación entre la depresión y la salud?» Primero, conviértala en una descripción más informativa que especifique un diseño del estudio, una variable predictiva, una variable de respuesta y una población. Después, discuta si esta pregunta de la investigación y este diseño que ha elegido cumplen los criterios FINER (Factible, Interesante, Novedoso, Ético y Relevante). Reformule la pregunta y el diseño para resolver cualquier problema en el cumplimiento de estos criterios.

- Considere la pregunta de la investigación: «¿El paracetamol produce asma?» Retroceda hasta el año 2000, cuando estaba empezando a plantearse esta pregunta, y ofrezca descripciones en una frase de dos estudios observacionales y un ensayo clínico en los que se abordó progresivamente esta pregunta de la investigación. Asegúrese de que cada una de las frases especifique el diseño del estudio, la variable predictiva, la variable de respuesta y la población. Después, para cada una, considere si esta pregunta de la investigación y el diseño que ha elegido cumplen los criterios FINER (Factible, Interesante, Novedoso, Ético y Relevante).

- Utilice las ideas de este capítulo y sus propios intereses para concebir una pregunta de la investigación y elabore un resumen en una página de un estudio que podría realizar. ¿Cumple los criterios FINER? Discuta diferentes diseños, muestras de la población y variables con un compañero, intentando optimizar la naturaleza FINER de su estudio.

Capítulo 3 Elección de los participantes del estudio: especificación, muestreo e inclusión

- Un investigador está interesado en la siguiente pregunta de la investigación: «¿Cuáles son los factores que hacen que las personas empiecen a fumar?» Decide utilizar una muestra transversal de estudiantes de instituto, invita a alumnos de segundo de bachillerato de su instituto local a participar, y estudia a los que se prestan voluntarios.

    - Comente la idoneidad de esta muestra para la población objetivo de interés.

    - Suponga que el investigador decide evitar el sesgo asociado a la elección de voluntarios diseñando una muestra aleatoria al 25 % de todos los alumnos de segundo de bachillerato, y que la muestra real está constituida por un 70 % de mujeres. Si se sabe que participaron aproximadamente el mismo número de hombres que de mujeres de segundo de bachillerato, la desproporción en la distribución por sexos representa un error en la obtención de la muestra. ¿Podría haberse producido por error aleatorio, error sistemático o ambos? Explique su respuesta.

- Un investigador está considerando diseños para realizar una encuesta entre asistentes habituales a conciertos de rock para determinar su actitud en relación con llevar tapones en los oídos durante los conciertos para proteger su audición. Diga el nombre de los siguientes esquemas de muestreo para seleccionar a participantes que rellenen un breve cuestionario, comentando la viabilidad y si los resultados podrán generalizarse a todas las personas que asisten a conciertos de rock.

    - Cuando cada cliente entra en el teatro, se le pide que tire un dado virtual (en el teléfono móvil del investigador). Se invita a todos los asistentes que saquen un 6 a que rellenen el cuestionario.

    - Cuando cada cliente entra en el teatro, se le pide que tire un dado virtual. Se selecciona a los varones que saquen un 1 y a las mujeres que saquen un número par.

    - Se sabe que las entradas para el concierto están numeradas y que se venden en la taquilla en orden sucesivo, y se selecciona a todos los clientes cuyo número de entrada termine en 1.

    - Una vez que todos estén sentados, se escogen cinco filas al azar, sacándolas de una baraja de cartas en las que cada carta corresponde a una fila del teatro. Se invita a todos los clientes que estén en esas filas.

    - Se invita a los 100 primeros clientes que entren al teatro.

    - Algunas entradas se han vendido por correo y otras en la taquilla justo antes del concierto. Siempre que haya tres o más personas esperando en la cola para comprar entradas en la taquilla, se invita a la última persona que esté en la cola (la que dispone de más tiempo).

    - Cuando las personas empiezan a marcharse tras el concierto, se invita a todos los que parecen dispuestos y capaces de rellenar el cuestionario.

- Edwards y cols. (Edwards y cols., *N Engl J Med* 2013;368:633-643) describieron la magnitud de la infección producida por el metaneumovirus humano (MPVH) en niños < 5 años de edad. Los participantes eran niños de condados de alrededor de Cincinnati, Nashville y Rochester, NY, durante los meses de noviembre a mayo, de 2003 a 2009, que solicitaron asistencia médica

por una infección respiratoria aguda o por fiebre. Se incluyó a los pacientes ingresados que dieron su consentimiento de domingo a jueves, a los pacientes ambulatorios 1 o 2 días a la semana, y a los pacientes del servicio de urgencias de 1 a 4 días a la semana. Los autores combinaron la proporción de niños con una prueba positiva en cada uno de los centros con datos nacionales (de los estudios *National Ambulatory Medical Care Survey* y *National Hospital Ambulatory Care Survey*) para determinar la frecuencia en la población de las visitas por enfermedad respiratoria aguda o fiebre a fin de estimar la magnitud total del MPVH en los Estados Unidos. Estimaron que el MPVH era responsable de 55 visitas a la consulta y 13 visitas al servicio de urgencias por cada 1 000 niños cada año.

- ¿Cuál fue la población objetivo en este ensayo?
- ¿Cuál era la población accesible, y hasta qué punto era adecuada para generalizar a la población objetivo?
- ¿Cuál fue el esquema de muestreo, y en qué medida era adecuado para generalizar a la población accesible?
- Describa en términos generales cómo habría que tener en consideración el esquema de muestreo para el cálculo de los intervalos de confianza de la incidencia de MPVH que se calculan.

Capítulo 4 Planificación de las mediciones: precisión, exactitud y validez

- Clasifique las siguientes variables como: dicotómica, nominal, ordinal, continua o numérica discreta. ¿Puede modificarse alguna de ellas para aumentar la potencia? ¿Cómo?
  - Antecedente de infarto de miocardio (presente/ausente)
  - Edad
  - Formación (título universitario o más/menos de título universitario)
  - Formación (máximo curso de escolarización)
  - Raza
  - Número de bebidas alcohólicas al día
  - Depresión (ausente, leve, moderada, grave)
  - Porcentaje de oclusión de las arterias coronarias
  - Color del pelo
  - Obesidad (IMC $\geq$ 30)/ausencia de obesidad (IMC $<$ 30)
- Un investigador está interesado en la pregunta de la investigación: «¿La ingesta de zumo de fruta a los 6 meses de edad predice el peso corporal con 1 año de edad?» Planifica un estudio de cohortes prospectivo, midiendo el peso corporal con una báscula para lactantes. Se observan algunos problemas durante las pruebas preliminares. ¿Se deben estos problemas a la falta de exactitud, a la falta de precisión o a ambas cosas? ¿Se debe, fundamentalmente, el problema a la variabilidad del observador, del participante o del instrumento, y qué podría hacerse?
  - Durante la calibración de la báscula, una pesa de referencia de 10 kg pesa 10,2 kg.
  - La báscula parece proporcionar resultados variables, pero, pesando 20 veces la pesa de referencia de 10 kg, se obtiene una media de 10,01 kg 6 0,2 (desviación típica) kg.
  - Algunos bebés se asustan, y cuando intentan bajarse de la báscula, el observador los sostiene para finalizar la medición.
  - Algunos bebés están intranquilos, y el indicador de la báscula se balancea arriba y abajo violentamente.
  - Algunos de los bebés llegan para la exploración inmediatamente después de haber comido, mientras que otros tienen hambre; algunos tienen los pañales mojados.
- El investigador tiene interés en estudiar el efecto de las limitaciones de las horas de trabajo de los residentes en los residentes

quirúrgicos. Un aspecto que desea abordar es el síndrome de desgaste profesional, y tiene previsto evaluarlo con dos preguntas (respondidas en una escala de 7 puntos) procedentes de un cuestionario más extenso: *a)* «¿con qué frecuencia siente usted que su trabajo le produce desgaste profesional»?, y *b)* «¿con qué frecuencia nota que se ha vuelto más duro con los demás desde que comenzó la residencia?»

El investigador se dispone a evaluar la validez de esas preguntas para medir el síndrome de desgaste profesional. Para cada una de las siguientes descripciones, nombre el tipo de validez que se evalúa:

- Los residentes con mayores puntuaciones de desgaste tenían más probabilidad de abandonar la residencia en el año siguiente.
- Estos aspectos parecen preguntas razonables que se deben plantear para estudiar el síndrome de desgaste profesional.
- Las puntuaciones de desgaste profesional aumentan durante las rotaciones más difíciles y disminuyen en vacaciones.
- En un estudio previo de más de 10 000 estudiantes de medicina, residentes y médicos en ejercicio, se demostró que estas dos preguntas detectaban casi por completo los dominios de agotamiento emocional y despersonalización medidos con el cuestionario de agotamiento de Maslach *(Maslach Burnout Inventory),* utilizado ampliamente, pero mucho más largo (West y cols., J Gen Intern Med 2009;24:1318-1321).

Capítulo 5 Preparación para el cálculo del tamaño de la muestra: hipótesis y principios subyacentes

- Defina los conceptos que aparecen en letra roja.

  Un investigador está interesado en diseñar un estudio con un tamaño de la muestra suficiente para determinar si el índice de masa corporal se asocia a cáncer de estómago en mujeres de entre 50 y 75 años de edad. Está planificando un estudio de casos y testigos con el mismo número de casos y de testigos. La hipótesis nula es que no existe ninguna diferencia en cuanto al índice de masa corporal entre los casos con cáncer de estómago y los testigos; ha elegido una hipótesis alternativa bilateral. Desearía tener una potencia de 0,80, con un nivel de significación estadística (α) de 0,05, para poder detectar una magnitud del efecto de una diferencia en el índice de masa corporal de 1 kg/m^2 entre los casos y los testigos. La revisión de la bibliografía indica que la variabilidad del índice de masa corporal en mujeres corresponde a una desviación típica de 2,5 kg/m^2.

- ¿Cuál de los siguientes puede ser un ejemplo de un error de tipo I? ¿Y un error de tipo II? ¿Ninguno?
  - En un ensayo aleatorizado se observa que los participantes tratados con un nuevo analgésico tienen mayores reducciones medias de las puntuaciones de dolor que los que reciben el placebo ($p = 0,03$).
  - En un estudio de diez años, comunica que 110 participantes que fuman no presentan mayor incidencia de cáncer de pulmón que 294 no fumadores (p = 0,31).
  - Un investigador llega a la conclusión de que «nuestro estudio es el primero que observa que el consumo de alcohol reduce el riesgo de sufrir diabetes en varones menores de 50 años de edad ($p < 0,05$)».

Capítulo 6 Cálculo de la potencia y del tamaño de la muestra: aplicaciones y ejemplos

- Revise el ejercicio 1 del capítulo 5. Determine cuántos casos de cáncer de estómago se necesitarían para el estudio. ¿Qué pasaría si el investigador deseara una potencia de 0,90? ¿O un nivel de significación estadística de 0,01?

  Crédito extra: Suponga que el investigador solo tiene acceso a 60 casos. ¿Qué podría hacer?

- La fuerza muscular disminuye al aumentar la edad. Las pruebas preliminares sugieren que parte de esta pérdida de fuerza muscular podría deberse a una carencia progresiva de deshidroepiandrosterona (DHEA). Los investigadores planean un ensayo aleatorizado para administrar DHEA o un placebo idéntico, durante 6 meses, a participantes ancianos, y medir después la fuerza muscular. En estudios anteriores se ha comunicado una fuerza media de prensión en personas ancianas de 20 kg, con una desviación típica de 8 kg. Suponiendo un valor de α (bilateral) = 0,05 y un valor de $\beta = 0,10$, ¿cuántos participantes se necesitarían para demostrar una diferencia del 10 % o mayor entre la fuerza del grupo tratado y la del grupo con placebo? ¿Cuántos participantes se necesitarían si $\beta = 0,20$?

- En el ejercicio 2, los cálculos del tamaño de la muestra indicaban que se necesitaban más participantes de los que podían incluirse. Un compañero señala que las personas ancianas tienen grandes diferencias de fuerza muscular. Esto explica gran parte de la

variabilidad de la fuerza medida tras el tratamiento, y podría no dejar ver el efecto de este. Le sugiere que mida la fuerza antes y después del tratamiento, usando el cambio de la fuerza como variable de respuesta. En un pequeño estudio piloto se ve que la desviación típica del cambio de la fuerza durante un período de 6 meses es de solo 2 kg. ¿Cuántos participantes se necesitarían por grupo si se usa este diseño, suponiendo que α (bilateral) = 0,05 y que β = 0,10?

- Un investigador sospecha que ser zurdo es más frecuente en los alumnos de tercer curso disléxicos que en los que no son disléxicos. En estudios anteriores se vio que alrededor del 10 % de las personas son zurdas y que la dislexia es infrecuente. Se planifica un estudio de casos y testigos que seleccionará a todos los estudiantes disléxicos de un distrito escolar como casos, con un mismo número de estudiantes no disléxicos seleccionados aleatoriamente como testigos. ¿Qué tamaño de la muestra se necesitaría para mostrar que la razón de posibilidades de dislexia es de 2,0 en los estudiantes zurdos en comparación con los estudiantes no zurdos? Suponga que α (bilateral) = 0,05 y que β = 0,20.

- Un investigador pretende determinar el CI medio de los estudiantes de medicina de su centro, con un IC del 99 % de ± 3 puntos. En un pequeño estudio piloto se ve que las puntuaciones del CI en los estudiantes de medicina oscilan desde alrededor de 110 hasta 150. ¿Cuál es, aproximadamente, el tamaño de la muestra que se necesita?

Capítulo 7 Diseño de estudios transversales y de cohortes

- La pregunta de la investigación es: «¿La carencia de vitamina B_{12} causa fracturas de cadera en el anciano?»
    - Resuma brevemente un plan del estudio para abordar esta pregunta de la investigación con un estudio de cohortes adecuado.
    - Un método alternativo sería obtener una muestra de una población clínica geriátrica y comparar la carencia de vitamina B_{12} en los que han tenido una fractura de cadera previamente con las concentraciones en los que no la han tenido. En comparación con este método de *transversal*, mencione, al menos, una ventaja y un inconveniente de su estudio de cohortes prospectivo.
    - ¿Podría diseñarse el estudio de cohortes como un estudio retrospectivo? De ser así, ¿cómo afectaría a estas ventajas o inconvenientes?

- Sung y cols. (Sung y cols., *Am J Obstet Gynecol* 2009 May;200(5):557.e1-5) evaluaron la asociación en situación inicial entre la frecuencia de la incontinencia urinaria y los síntomas depresivos en 338 mujeres con sobrepeso u obesas de al menos 30 años de edad incluidas en el ensayo clínico PRIDE (*Program to Reduce Incontinence by Diet and Exercise*). Observaron que las mujeres con síntomas depresivos (n = 101) referían un mayor número medio de episodios de incontinencia urinaria por semana que las mujeres sin síntomas depresivos (28 y 23, respectivamente; p = 0,005).
    - ¿Qué tipo de estudio es este?
    - Una posible explicación de este fenómeno es que la depresión aumenta la frecuencia de la incontinencia urinaria. ¿Cuáles son algunas otras explicaciones de esta asociación, y cómo podría diferenciarlas mediante modificaciones del diseño del estudio?

Capítulo 8 Diseño de estudios de casos y testigos

- La pregunta de la investigación es: «¿Hasta qué punto un antecedente familiar de cáncer de ovario aumenta el riesgo de sufrir este tipo de cáncer?» El investigador planifica un estudio de casos y testigos para responder a esta pregunta.
    - ¿Cómo debe escoger los casos?
    - ¿Cómo debe escoger los testigos?
    - Comente las posibles fuentes de sesgo en el muestreo de los casos y de los testigos.
    - ¿Cómo mediría el «antecedente familiar de cáncer de ovario» como variable predictiva de interés? Comente las fuentes de sesgo en esta medición.
    - ¿Qué medida de asociación usaría, y qué prueba de significación estadística?
    - ¿Cree que el método de casos y testigos es el adecuado para esta pregunta de la investigación? Comente las ventajas y los

inconvenientes del diseño de casos y testigos con respecto a otras posibilidades para esta pregunta.

- El investigador desea estudiar la relación entre jugar con videojuegos de carreras de coches y el riesgo de verse implicado en un accidente automovilístico real (como conductor).
    - Suponga que la exposición de interés es el efecto a largo plazo del uso habitual de estos juegos. ¿Cómo seleccionaría los casos y los testigos y mediría la exposición en un estudio de casos y testigos para abordar esta pregunta?
    - Ahora imagine que la exposición de interés es si el uso de dichos juegos en la hora inmediatamente anterior a conducir aumenta el riesgo a corto plazo. ¿Cuál es el diseño para estudios de efectos a corto plazo de exposiciones intermitentes? Describa cómo se realizaría un estudio de este tipo para esta pregunta de la investigación.

Capítulo 9 Mejora de la inferencia causal en estudios de observación

- El investigador emprende un estudio de casos y testigos para la pregunta de la investigación «¿El consumo de más frutas y verduras reduce el riesgo de sufrir cardiopatía isquémica (CI) en el anciano?» Suponga que su estudio demuestra que las personas del grupo testigo comunican un mayor consumo de frutas y verduras que las personas con CI.

  ¿Cuáles son las posibles explicaciones para esta asociación inversa entre el consumo de frutas y verduras y la CI? Preste mucha atención a la posibilidad de que el ejercicio pueda introducir confusión en la asociación entre el consumo de frutas y verduras y la CI (si las personas que consumen más frutas y verduras también realizan más ejercicio, y esta es la causa de su menor incidencia de CI). ¿Qué métodos podría usar para contar con el ejercicio como una posible variable de confusión, y cuáles son las ventajas y los inconvenientes de cada plan?

- En un estudio de la red PROS (*Pediatric Research in Office Settings*) de pediatras se encontró que en lactantes pequeños (< 3 meses) llevados a su pediatra por fiebre, los niños no circuncidados tenían aproximadamente 10 veces más riesgo de infección del aparato urinario que los niños circuncidados (Newman y cols., *Arch Pediatr Adolesc Med* 2002 Jan;156(1):44-54), asociación que se ha visto en numerosos estudios. Lo interesante es que los niños no circuncidados de ese estudio parecían tener *menor* riesgo de infecciones óticas (cociente de riesgos instantáneos = 0,77; $p = 0,08$). Explique cómo solo la inclusión de lactantes con fiebre en este estudio podría introducir una asociación entre la circuncisión y las infecciones óticas que no está presente en la población general de lactantes pequeños.

- En el ejercicio 1 del capítulo 2 le pedimos que propusiera estudios para abordar la pregunta de si el paracetamol produce asma. Un mecanismo propuesto de esta asociación es la depleción inducida por el paracetamol de glutatión, que protege a los pulmones de las agresiones oxidativas que pueden producir inflamación. Describa brevemente cómo aprovecharía la variación de los genotipos de los antioxidantes maternos para mejorar la inferencia de que una asociación entre el consumo materno de paracetamol y el asma en la descendencia es causal.

Capítulo 10 Diseño de un estudio aleatorizado con enmascaramiento

- Un extracto herbal, la huperzina, se ha utilizado en China como remedio frente a la demencia, y en estudios preliminares realizados en animales y seres humanos se han visto resultados prometedores. El investigador desearía probar si este nuevo tratamiento podría disminuir la progresión de la enfermedad de Alzheimer. Se ha observado que la concentración plasmática de β-a-miloide (Abeta) (1-40) es un marcador biológico de la enfermedad de Alzheimer: las concentraciones elevadas se asocian a un riesgo significativamente mayor de sufrir demencia, y las concentraciones de Abeta (1-40) aumentan con la progresión de la demencia. Al planificar un ensayo para probar la eficacia de la huperzina en la prevención de la demencia en pacientes ancianos con una alteración cognitiva leve, el investigador considera dos posibles mediciones de la respuesta: el cambio de las concentraciones de Abeta (1-40) y la incidencia de un diagnóstico clínico de demencia.
    - Mencione una ventaja y un inconveniente del uso de Abeta (1-40) como respuesta principal en su ensayo.
    - Mencione una ventaja y un inconveniente del uso del nuevo diagnóstico de demencia como la respuesta principal en su estudio.

- Se está planificando un ensayo clínico relativamente extenso (> 200 personas por grupo) de huperzina. El principal objetivo es evaluar si este extracto herbal disminuye la incidencia de diagnóstico clínico de demencia en varones y mujeres ancianos con una alteración cognitiva leve.
    - Se espera que la huperzina cause, en ocasiones, síntomas gastrointestinales, como diarrea, náuseas y vómitos. Describa un plan para valorar los efectos adversos de este nuevo tratamiento sobre otros síntomas o enfermedades aparte de la

demencia.

- Describa un plan general para la recopilación de los datos iniciales: ¿qué tipos de información debe recopilar?

- Las personas portadoras de un alelo Apoε4 tienen mayor riesgo de sufrir demencia. Mencione una razón a favor y otra en contra de usar la aleatorización estratificada por conglomerados bloques en lugar de la aleatorización habitual para asegurar un equilibrio de las personas con el genotipo Apoε4 en los grupos de tratamiento y testigo.

Capítulo 11 Otros diseños de ensayos clínicos y problemas en su realización

La finasterida tópica es moderadamente eficaz en el tratamiento de la calvicie de patrón masculino, y está aprobada por la Food and Drug Administration (FDA) de Estados Unidos para tratar esta afección. Se ha observado que las estatinas aumentan el crecimiento del pelo en los roedores, y que actúan por una vía diferente a la de la finasterida. Imagine que una compañía que empieza desea conseguir la aprobación de la FDA para comercializar una nueva estatina tópica (HairStat) para el tratamiento de la calvicie de patrón masculino.

- Describa un ensayo de fase I de HairStat para la calvicie de patrón masculino. ¿Cuáles deben ser los grupos de tratamiento? ¿Qué tipo de respuestas cabe esperar?

- La compañía desea comparar la eficacia de HairStat con la de la finasterida. Diga, al menos, una ventaja y un inconveniente de los siguientes métodos para probar la eficacia relativa de la finasterida y la estatina tópica.

  - Aleatorizar a varones calvos a recibir la finasterida o la estatina tópica.

  - En un diseño factorial, asignar aleatoriamente a varones a recibir: *1)* finasterida y HairStat; *2)* finasterida y placebo de HairStat; *3)* placebo de finasterida y HairStat, y *4)* doble placebo.

- Imagine que la compañía planea un estudio con placebo de 1 año controlado de duración de HairStat para el tratamiento de la calvicie. La respuesta es el cambio de la puntuación de la cantidad de pelo en las fotografías de la región con calvicie que está recibiendo tratamiento. Las consultas de seguimiento (con fotografías) se programan cada 3 meses. Resuma un plan (con al menos dos elementos) para estimular el cumplimiento del tratamiento en estudio y de las consultas para evaluar la respuesta.

- Durante el estudio, el 20 % de los varones del ensayo no acudió a la visita de seguimiento a los 3 meses, y el 40 % abandonó al cabo de 1 año. Algunos lo hicieron debido a la aparición de una erupción en el cuero cabelludo. Mencione un inconveniente y una ventaja de analizar el efecto del tratamiento sobre el crecimiento del cabello mediante un método estricto de intención de tratar.

- En el análisis de intención de tratar, HairStat aumentó el crecimiento capilar (determinado por evaluadores de la respuesta que desconocían la asignación al tratamiento, basándose en la comparación de las fotografías inicial y transcurrido 1 año) un 20 % más que el placebo ($p = 0{,}06$). Los análisis posteriores mostraron que HairStat aumentó el crecimiento del pelo un 45 % más que el placebo en los varones de menos de 40 años de edad ($p = 0{,}01$ en ese subgrupo). ¿Qué problemas presenta la conclusión de la compañía de que HairStat es eficaz en el tratamiento de la calvicie de varones de menos de 40 años de edad?

Capítulo 12 Diseño de estudios de pruebas médicas

- Está usted interesado en estudiar la utilidad de la velocidad de sedimentación globular (VSG) como prueba para diagnosticar enfermedad inflamatoria pélvica (EIP) en mujeres con dolor abdominal.

  - Para hacerlo, necesitará reunir grupos de mujeres que tengan y que no tengan EIP. ¿Cuál sería el mejor modo de muestrear estas mujeres?

  - ¿Cómo podrían estar sesgados los resultados si utiliza el diagnóstico final de EIP como método de referencia y quienes asignan ese diagnóstico conocen la VSG?

  - Observa que la sensibilidad de una VSG de al menos 20 mm/h es del 90 %, pero la especificidad es solo del 50 %. Por otro lado, la sensibilidad de una VSG de al menos 50 mm/h es solo del 75 %, pero la especificidad es del 85 %. ¿Qué valor de corte utilizaría usted para definir una VSG anormal?

- Está interesado en estudiar el rendimiento diagnóstico de la tomografía computarizada (TC) craneal en niños que acuden al servicio de urgencias (SU) con lesiones craneales. Utiliza una base de datos del departamento de radiología para encontrar registros de todas las TC realizadas a pacientes de menos de 18 años y solicitadas por el SU por un traumatismo craneal. A continuación revisa los registros del SU de todos aquellos en los que la TC fue anormal, para determinar si la alteración podía

haberse pronosticado a partir de la exploración física.

- De 200 TC, había lesiones intracraneales en 10. Sin embargo, usted determina que en 8 de 10 había una exploración neurológica focal o una alteración del estado mental. Ya que solo dos pacientes tenían TC anormales que podían no haberse pronosticado a partir de la exploración física, llega a la conclusión de que el rendimiento de lesiones intracraneales «inesperadas» es solo de 2 de 200 (1 %) en este caso. ¿Qué falla en esta conclusión?

- ¿Qué falla con el uso de todas las lesiones intracraneales identificadas en la TC como la variable de respuesta en este estudio de rendimiento diagnóstico?

- ¿Cuáles serían algunas de las ventajas de estudiar los efectos de la TC en la toma de decisiones clínicas, en lugar de solo en el rendimiento diagnóstico?

- Ahora desea estudiar la sensibilidad y la especificidad de los hallazgos neurológicos de focalidad para predecir las lesiones intracraneales. (Debido al pequeño tamaño de la muestra de pacientes con lesiones intracraneales, aumenta su tamaño ampliando el estudio a otros SU.) Un problema que tiene estudiar hallazgos neurológicos focales es que es más probable que se realice una TC a los niños que los presentan que a los que no los tienen. Explique cómo y por qué afectará esto a la sensibilidad y la especificidad de esos hallazgos si:

  - Se incluye en el estudio solo a los niños a los que se realizó una TC.

  - Se incluye a niños elegibles con lesiones craneales a los que no se realizó una TC, y se supone que no han tenido una lesión craneal si se recuperaron sin necesidad de una intervención neuroquirúrgica.

Capítulo 13 Investigación utilizando datos existentes

- La pregunta de la investigación es: «¿Tienen los latinos estadounidenses mayor frecuencia de colecistopatía que los caucásicos, los afroamericanos y las personas de origen asiático?» ¿Qué bases de datos existentes podrían facilitarle la determinación de las incidencias de colecistopatía específicas de raza, edad y sexo con un bajo coste en cuanto a tiempo y dinero?

- Un becario de investigación está interesado en la pregunta de si la insuficiencia renal leve o moderada aumenta el riesgo de sufrir episodios de cardiopatía isquémica y muerte. Debido al coste y la dificultad que supone realizar un estudio para generar datos primarios, buscó una base de datos que contuviera las variables que necesitaba para responder su pregunta de la investigación. Observó que el estudio *Cardiovascular Health Study* (CHS), un extenso estudio de cohortes, multicéntrico y financiado por el NIH, de factores pronósticos de enfermedad cardiovascular en hombres y mujeres ancianos, proporcionaba todas las variables necesarias para su análisis. Su mentor consiguió presentarle a uno de los investigadores principales del estudio, y este le ayudó a preparar y enviar una propuesta para su análisis, que fue aprobada por el comité directivo del estudio CHS.

  - ¿Qué ventajas presenta este método para estudiar esta pregunta?

  - ¿Cuáles son sus inconvenientes?

- Un investigador está interesado en si los efectos del tratamiento con estrógenos después de la menopausia o con moduladores selectivos de los receptores estrogénicos (SERM) varían dependiendo de las concentraciones de estrógenos endógenos. ¿Cómo podría este investigador responder a la pregunta usando un estudio auxiliar?

Capítulo 14 Abordaje de cuestiones éticas

- La pregunta de la investigación es identificar genes que estén asociados con mayor riesgo de sufrir diabetes mellitus de tipo 2. El investigador observa que puede disponer de muestras de sangre congeladas y datos clínicos de un gran estudio de cohortes prospectivo sobre factores de riesgo de cardiopatía isquémica que ya se ha realizado. En ese estudio se recopilaron datos basales sobre dieta, ejercicio, características clínicas y concentraciones de colesterol y hemoglobina A1c. Dispone de datos de seguimiento de criterios de valoración coronarios y aparición de diabetes. El estudio propuesto realizará secuenciación del ADN de los participantes; no se precisan nuevas muestras de sangre.

  - ¿Puede realizarse el estudio propuesto con el consentimiento informado original que se obtuvo para el estudio de cohortes?

  - Si el consentimiento original no daba autorización para el estudio propuesto, ¿cómo puede llevarse a cabo el estudio propuesto?

- Al diseñar nuevos estudios en los que se obtengan muestras de sangre, ¿cómo pueden los investigadores planificar la autorización del uso de sus datos y muestras en estudios futuros?

- El investigador planifica un estudio de fase II, controlado y aleatorizado, de un nuevo fármaco antineoplásico que ha dado resultados prometedores en el cáncer de colon. Para reducir el tamaño de la muestra, desearía realizar un estudio controlado con placebo, en lugar de compararlo con el tratamiento actual.

  - ¿Qué cuestiones éticas surgen acerca de un testigo con placebo en esta situación?

  - ¿Es posible realizar un estudio controlado con placebo de un modo éticamente aceptable?

- El investigador planifica un estudio en preparación para un futuro ensayo de una vacuna frente al VIH. Los objetivos son determinar: *1)* si es posible reunir una cohorte de participantes con una incidencia elevada de seroconversión al VIH a pesar del mejor asesoramiento sobre la prevención frente al VIH, y *2)* si la tasa de seguimiento en la cohorte será lo suficientemente elevada como para realizar un ensayo de la vacuna. Los participantes serán personas con riesgo elevado de infección por el VIH, entre ellos consumidores de drogas por vía intravenosa, personas que comercian con el sexo y otras con múltiples parejas sexuales. La mayoría de los participantes tendrá una formación escasa y unos conocimientos sanitarios inadecuados. El estudio será un estudio de cohortes de observación, en el que el seguimiento de los participantes será de 2 años, para determinar las tasas de seroconversión y de seguimiento.

  - ¿Qué exigen las regulaciones federales que se comunique a los participantes como parte del consentimiento informado?

  - ¿Qué pasos pueden adoptarse para garantizar que el consentimiento sea realmente informado a este respecto?

  - ¿Cuál es la responsabilidad del investigador durante este estudio de observación para reducir el riesgo de infección por el VIH en estos pacientes?

Capítulo 15 Diseño de cuestionarios, entrevistas y encuestas por Internet

- Como parte de un estudio sobre el alcohol y la fuerza muscular, un investigador planifica usar el siguiente punto para un cuestionario de autorrespuesta para determinar el consumo actual de alcohol:

  «¿Cuántos vasos de cerveza, vino o licor bebe cada día?»

  ○ 0

  ○ 1-2

  ○ 3-4

  ○ 5-6

  ○ 7-8

  Describa brevemente, al menos, dos problemas que presente esta pregunta.

- Escriba una serie corta de preguntas para un cuestionario de autorrespuesta que valoren mejor el consumo actual de alcohol.

- Comente las ventajas y los inconvenientes de un cuestionario de autorrespuesta frente a una entrevista estructurada para valorar la conducta sexual de riesgo.

Capítulo 16 Gestión de los datos

- Consulte las seis primeras preguntas del cuestionario de muestra acerca del tabaquismo del apéndice 15. Tiene respuestas de tres participantes del estudio:

ID del participante	**Descripción de la historia de tabaquismo**

1 001	Empezó a fumar a los 17 años y ha seguido fumando un promedio de 30 cigarrillos/día desde entonces
1 002	Empezó a fumar a los 21 años y ha fumado 20 cigarrillos/día hasta que lo dejó, hace 3 años, a la edad de 45
1 003	Fumó algunos cigarrillos (< 100) en los años de enseñanza secundaria

Cree una tabla de datos que contenga las respuestas de estos participantes a las seis primeras preguntas del apéndice 15. La tabla debe tener tres filas (una para cada participante) y siete columnas (una para el ID del participante, y una para cada una de las seis preguntas).

- El estudio PHTSE (*Pre-Hospital Treatment of Status Epilepticus*) (Lowenstein y cols., Control Clin Trials 2001;22:290-309; Alldredge y cols., N Engl J Med 2001;345:631-637) fue un ensayo para el tratamiento del estado epiléptico prehospitalario. El criterio de valoración principal fue la desaparición de las convulsiones al llegar al hospital. Para incluir participantes, los paramédicos contactaron con médicos del hospital por radio. A continuación se muestran los formularios para la recopilación de datos por médicos del hospital de dos pacientes incluidos:

PHTSE

Formulario de recopilación de datos por médicos del hospital

ID del participante del PHTSE:

Administración del fármaco en estudio: *189*

Equipo del fármaco del estudio n.º: *A322*

Fecha y hora de administración: *3 / 12 / 94* *17 : 39* (Use reloj de 24 h)

Evaluación del transporte

Cesaron las convulsiones

Hora a la que cesaron las convulsiones: *17 : 44* (Use reloj de 24 h)

Valoración final

Hora de llegada al SU del hospital de destino: *17 : 48* (Use reloj de 24 h)

Al llegar al hospital de destino:

[X] 1 Continuaba la actividad convulsiva (convulsiones tonicoclónicas activas)

[ ] 0 Cesó la actividad convulsiva

ECG verbal

- [ ] 1 Sin respuesta verbal
- [ ] 2 Habla incomprensible
- [ ] 3 Habla inadecuada
- [ ] 4 Habla confusa
- [ ] 5 Orientado

¡¡PÁGINA CORRESPONDIENTE AL TURNO DE GUARDIA DEL PERSONAL DEL ESTUDIO PHTSE!!

Herramientas de imágenes

PHTSE

Formulario de recopilación de datos por médicos del hospital

ID del participante del PHTSE:

Administración del fármaco en estudio: *410*

Equipo del fármaco del estudio n.º: *B536*

Fecha y hora de administración: *12 / 01 / 98* *01 : 35* (Use reloj de 24 h)

Evaluación del transporte

[X] Cesaron las convulsiones

Hora a la que cesaron las convulsiones: *01 : 39* (Use reloj de 24 h)

Valoración final

Hora de llegada al SU del hospital de destino: *01 : 53* (Use reloj de 24 h)

Al llegar al hospital de destino:

[ ] 1 Continuaba la actividad convulsiva (convulsiones tonicoclónicas activas)

[X] 0 Cesó la actividad convulsiva

ECG verbal

[ ] 1 Sin respuesta verbal
[ ] 2 Habla incomprensible
[ ] 3 Habla inadecuada
[X] 4 Habla confusa
[ ] 5 Orientado

¡¡PÁGINA CORRESPONDIENTE AL TURNO DE GUARDIA DEL PERSONAL DEL ESTUDIO PHTSE!!

Herramientas de imágenes

- Disponga los datos de estos dos formularios de recogida de datos en una tabla de datos con dos filas.
- Cree un diccionario de los datos de nueve campos para la tabla de datos del ejercicio 2a.
- Los formularios en papel para la recogida de datos fueron rellenados por médicos del hospital muy ocupados, a los que llamaron desde el servicio de urgencias a una sala de radio. ¿Cuáles son las ventajas y los inconvenientes de usar un formulario informático en pantalla en lugar de un formulario en papel? Si usted diseñara el estudio, ¿qué formulario usaría?

- Los formularios de recogida de datos del ejercicio 2 incluyen una pregunta sobre si la actividad convulsiva seguía al llegar al hospital receptor (que era la variable de respuesta principal del estudio). A esta pregunta se le asignó el nombre de campo *ActConLleHosp* y se codificó con 1 para el sí (la actividad convulsiva continuaba) y con 0 para el no (la actividad convulsiva cesó).

  Interprete los valores promedio para *ActConLleHosp* como se muestra a continuación:

ActConLleHosp

(1 = *Sí, las convulsiones seguían;* 0 = *No, las convulsiones habían cesado)*

	N	Promedio
Lorazepam	66	0,409
Diazepam	68	0,574
Placebo	7	0,789

Capítulo 17 Realización del estudio y control de calidad

- Un investigador realizó un estudio sobre la pregunta de la investigación: «¿Cuáles son los factores pronósticos de muerte tras la hospitalización por infarto de miocardio?» Los auxiliares de la investigación recopilaron datos detallados de las historias y realizaron entrevistas extensas a 120 pacientes hospitalizados a los que se siguió durante 1 año. Alrededor del 15 % de los pacientes falleció durante el período de seguimiento. Cuando estuvo completa la recogida de los datos, uno de los auxiliares los introdujo en un ordenador usando una hoja de cálculo. Cuando el investigador inició el análisis de datos, descubrió que faltaba el 10-20 % de algunas variables predictivas, y otras no parecían ser lógicas. Solo el 57 % de la muestra había acudido al seguimiento al cabo de 1 año, fecha desde la que había transcurrido actualmente más de 1 año para algunos pacientes. Se le ha llamado como consultor del proyecto.
    - ¿Qué puede hacer ahora el investigador para mejorar la calidad de sus datos?
    - Describa brevemente al menos tres formas mediante las que podría reducir los valores ausentes y los errores en su siguiente estudio.

Capítulo 18 Estudios comunitarios e internacionales

- El investigador decide estudiar las características y la evolución clínica de pacientes con dolor abdominal de etiología poco clara. Planifica incluir pacientes con dolor abdominal en los que no pueda identificarse una causa específica tras realizar una serie de pruebas habituales. Hay dos opciones para encontrar participantes para el estudio: *1)* la consulta de aparato digestivo de su centro médico universitario, o *2)* una red local de consultas de la comunidad. ¿Cuáles son las ventajas y los inconvenientes de cada método?
- Se ha destinado al investigador a trabajar con el Ministerio de Sanidad chino en un nuevo programa para evitar las enfermedades relacionadas con el tabaco en China. De las siguientes preguntas de investigación, ¿hasta qué punto hace falta para cada una de ellas investigación local en oposición a investigación realizada en otros lugares?
    - ¿Cuál es la frecuencia y la distribución del consumo de cigarrillos?
    - ¿Qué enfermedades están causadas por el tabaquismo?
    - ¿Qué estrategias son las más eficaces para animar a las personas a dejar de fumar?

Capítulo 19 Elaboración de una propuesta para la financiación de la investigación

- Busque en la página web de los NIH (http://grants.nih.gov/grants/oer.htm) para encontrar al menos tres tipos de subvenciones de serie R iniciadas por el investigador.
- Busque en la página web del Foundation Center (http://fdncenter.org/) fundaciones que podrían estar interesadas en su área de interés. Enumere, al menos, dos.

- Póngase en contacto con mentores y compañeros para encontrar un protocolo de investigación que aborde una pregunta de su área de interés y que haya sido financiado. Lea atentamente este protocolo.

Respuestas a los ejercicios
NA

Capítulo 1 Cómo empezar: anatomía y fisiología de las investigaciones clínicas

1a. Se trata de una inferencia de la validez interna (porque se refiere a las mujeres de este estudio) que probablemente sea válida. Sin embargo, podría ser inválida si algo distinto a la fórmula limitada temprana (FLT) produjera la diferencia de las tasas de lactancia materna (p. ej., si la intervención testigo afectará negativamente a la lactancia materna), si la lactancia materna referida por las propias madres no reflejara la lactancia materna real, o si la asociación no fuera causal (el valor de $p = 0{,}02$ no descarta que se haya producido por azar).

1b. Se trata de una inferencia de la validez externa (porque supone la generalización fuera del estudio) que puede ser válida. Sin embargo, además de las amenazas a la validez interna que se han señalado más arriba (y que también amenazan a la validez externa), es probable que las mujeres que tengan sus hijos en hospitales comunitarios y en otras partes del país puedan responder de manera diferente a la intervención, o que otros médicos que ofrezcan la FLT puedan aplicar la intervención de una forma diferente a como se realizó en el estudio original, o que los efectos beneficiosos pudieran no durar tanto como 6 meses.

1c. Es una inferencia de la validez externa que va mucho más allá de la población y la intervención que se estudiaron, y que probablemente no sea válida. Supone la generalización no solo a otras madres y recién nacidos en otras localizaciones, sino que también incluye a recién nacidos que no han perdido el 5 % del peso corporal; amplía la intervención desde la fórmula limitada temprana hasta aportar la fórmula sin limitación; y propone unos beneficios para la salud amplios e inespecíficos que, aunque son razonables, no se evaluaron en el estudio ELF.

2a. Es un estudio de cohortes de si ver combates de lucha en televisión predice las peleas posteriores en estudiantes de instituto de Winston-Salem.

2b. Es un estudio de casos y testigos de si la duración de la lactancia materna se asocia a reducción del riesgo de cáncer de ovario en mujeres chinas que han dado de mamar al menos a un hijo.

2c. Es un estudio transversal de la relación entre la ingesta de grasa saturada referida por los propios pacientes y la concentración de espermatozoides en hombres daneses a los que se evaluó para el servicio militar.

2d. Es un estudio aleatorizado de diseño abierto de si un ciclo corto de vancomicina, lavado intestinal e infusión duodenal de heces de donante mejora la tasa de curación a las 10 semanas en adultos con diarrea recurrente por *C. difficile* en comparación con un régimen estándar de vancomicina con o sin lavado intestinal.

Cada una de estas cuatro frases es una descripción breve que resume todo el estudio, señalando el diseño y los elementos principales de la pregunta de la investigación (variables principales y población). Por ejemplo, en el ejercicio 2a, el diseño es un estudio de cohortes, la variable predictiva es ver lucha en televisión, la variable de respuesta son las peleas, y la población son estudiantes de instituto de Winston-Salem.

Capítulo 2 Elaboración de la pregunta de la investigación y desarrollo del plan del estudio

- El proceso de pasar de la pregunta de la investigación al plan del estudio muchas veces es iterativo. Se podría comenzar con una pregunta como: «Un estudio transversal para determinar si la depresión está asociada al estado de salud en adultos jóvenes.» La posibilidad de que «depresión» y «estado de salud» estén relacionados parece Interesante y Relevante, pero la pregunta, tal y como se ha realizado, sigue siendo demasiado vaga como para poder juzgar si el estudio es Factible, Novedoso y Ético. ¿Cómo se medirán la depresión y el estado de salud, y en qué población? También será difícil establecer la causalidad en un estudio transversal: ¿la depresión empeora la salud o viceversa?

  Un diseño más específico que podría cumplir mejor los criterios FINER podría ser: «Un estudio de cohortes para determinar si la depresión en estudiantes de primer curso de universidad predice el número de visitas por enfermedad médica al servicio sanitario para estudiantes en el año siguiente.»

- En el caso de la asociación entre el paracetamol y el asma, la observación de que el uso de paracetamol y la prevalencia del asma han aumentado en todo el mundo (y la credibilidad biológica relacionada con la depleción de glutatión inducida por el paracetamol) llevó a que todos los estudios fueran Interesantes y Relevantes; a medida que se hagan más estudios, serán menos Novedosos.

Estudio n.º 1: estudio de casos y testigos para comparar la frecuencia autorreferida de consumo de paracetamol en adultos con síntomas de asma atendidos en consultas generales del sur de Londres (dos casos) con la frecuencia descrita por adultos seleccionados aleatoriamente sin dichos síntomas, procedentes de las mismas consultas de medicina general (los testigos). Los estudios de casos y testigos son, a menudo, una buena forma de comenzar a investigar posibles asociaciones (cap. 8). Este estudio era especialmente Factible, porque formaba parte de un estudio de casos y testigos poblacional más extenso que ya investigaba la importancia de los antioxidantes de la dieta en el asma. Las razones de posibilidades de asma aumentaron con la frecuencia del consumo de paracetamol, hasta 2,38 (IC del 95 %: 1,22-4,64) en los que lo consumían a diario; *p* para la tendencia = 0,0 002). El estudio fue Ético, porque se trató de un estudio observacional que no hizo que los pacientes tuvieran riesgo (Shaheen y cols., *Thorax* 2000;55:266-270).

Estudio n.º 2: estudio transversal multinacional de síntomas alérgicos descritos por los padres (asma, fiebre del heno y eccema) en niños de 6 a 7 años que incluyó preguntas sobre el uso de paracetamol en el año previo y el uso habitual por fiebre en el primer año después del nacimiento. Este estudio (en el que se incluyó a 205 487 niños de 6 a 7 años de edad de 71 centros de 31 países) no habría sido Factible si no hubiera formado parte del estudio más general *International Study of Asthma and Allergies in Childhood* (ISAAC). Esto ilustra la importancia de buscar datos existentes o estudios existentes cuando se investiga una nueva pregunta de la investigación (cap. 13). Los autores encontraron una sólida relación de dosis-respuesta entre el consumo actual de paracetamol y las sibilancias, y una razón de posibilidades de 1,46 (IC del 95 %: 1,36-1,56) para las sibilancias y una respuesta de «sí» a la pregunta: «en los primeros 12 meses de vida de su hijo, ¿le administró habitualmente paracetamol por fiebre?» (Beasley y cols., *Lancet* 2008;372:1 039-1 048).

Estudio n.º 3: estudio aleatorizado y con doble enmascaramiento del efecto del paracetamol (12 mg/kg) y del ibuprofeno (5 o 10 mg/kg) sobre los ingresos hospitalarios y las visitas ambulatorias por asma durante 4 semanas en niños febriles de 6 meses a 12 años de edad a los que se estaba tratando por asma en la inclusión. Un ensayo aleatorizado es, generalmente, el diseño menos Factible, debido al coste y la logística implicados. Además, a medida que se acumulan datos del posible efecto adverso del fármaco, los estudios aleatorizados para confirmarlo pasar a ser menos Éticos. En este caso, los investigadores realizaron un análisis retrospectivo de los datos del grupo de niños con asma del estudio *Boston University Fever Study,* un ensayo aleatorizado y con doble enmascaramiento cuya inclusión había finalizado en 1993. Encontraron que los niños asignados aleatoriamente a paracetamol tuvieron un riesgo un 59 % mayor de hospitalizaciones por asma (NS) y un riesgo un 79 % mayor de visitas ambulatorias por asma (RR = 1,79; IC del 95 %: 1,05, 2,94; *p* = 0,01) (Lesko y cols., *Pediatrics* 2002;109:E20).

Capítulo 3 Elección de los participantes del estudio: especificación, muestreo e inclusión

1a. Esta muestra de alumnos de bachillerato puede no ser adecuada para la pregunta de la investigación si los antecedentes de tabaquismo se producen a una edad anterior. Una población objetivo de mayor interés podrían ser los estudiantes al comienzo de la enseñanza secundaria. Además, la población accesible (los estudiantes de este instituto) podría no representar adecuadamente a la población objetivo, porque las causas del tabaquismo difieren en diversos contextos culturales, y el investigador debería extraer la muestra de varios institutos seleccionados aleatoriamente de toda la región. Lo que es más importante, es probable que el diseño del muestreo (petición de voluntarios) atraiga a estudiantes que no sean representativos de la población accesible por su conducta de tabaquismo.

1b. La muestra no representativa podría haberse debido al error aleatorio, aunque esto habría sido improbable, salvo que fuera una muestra muy pequeña. Si la muestra estuviera formada por 10 personas, una desproporción 7:3 se produciría con bastante frecuencia como consecuencia del azar; de hecho, la probabilidad de seleccionar al menos siete niñas de una clase grande que tiene el 50 % de niñas es de aproximadamente el 17 % (más otra probabilidad del 17 % de seleccionar al menos siete niños). Pero si el tamaño de la muestra fuera 100, la probabilidad de obtener una muestra de al menos 70 niñas sería menor del 0,01 %. Esto ilustra el hecho de que el investigador puede estimar la magnitud del componente aleatorio del error de muestreo una vez que ha obtenido la muestra, y puede reducir dicho error hasta cualquier nivel deseado aumentando el tamaño de la muestra.

La muestra no representativa también se podría deberse a un error sistemático. La gran proporción de mujeres se podría haber debido a diferentes tasas de participación de niños y niñas. Las estrategias para prevenir el sesgo de ausencia de respuesta incluyen el espectro de técnicas para mejorar la inclusión que se discute en el capítulo 3. La gran proporción de mujeres también podría representar un error técnico en la enumeración o la selección de los nombres a muestrear. Las estrategias para prevenir los errores incluyen el uso adecuado de estudio preliminar y procedimientos de control de calidad (cap. 17).

2a. Muestra aleatoria (probabilística). El principal problema en relación con la posibilidad de generalizar será la ausencia de respuesta; será importante que el cuestionario sea corto y ofrecer algún incentivo para rellenarlo. (El posible sesgo de ausencia de respuesta es problemático en todos los esquemas de muestreo que se discuten en esta pregunta.)

2b. Muestra aleatoria estratificada (probabilística), con un muestreo tres veces mayor de mujeres, tal vez porque el investigador previó que menos mujeres asistirían al concierto.

2c. Muestra sistemática (no probabilística). Aunque tal vez sea cómodo, este esquema de muestreo llevaría a una representación insuficiente de los dos miembros de las parejas. Además, al menos en teoría, el vendedor de la taquilla podría manipular qué asistentes reciben billetes que acaban en 1.

2d. Muestra de conglomerados (probabilística). Puede ser cómoda, aunque el agrupamiento se debe tener en consideración en los análisis, porque las personas que están sentadas en la misma fila pueden ser más similares entre sí que los asistentes al concierto seleccionados aleatoriamente. Esto podría ser un problema importante si la música fuera más intensa en algunas filas que en otras.

2e. Muestra consecutiva (no probabilística). Las muestras consecutivas habitualmente son una buena opción, pero las personas que llegan pronto a los conciertos pueden diferir de aquellas que llegan más tarde, por lo que sería preferible disponer de varias muestras consecutivas seleccionadas en momentos diferentes.

2f. Muestra de conveniencia (no probabilística). Este esquema no incluirá a las personas que hayan comprado las entradas por correo. Además, podría haber una representación excesiva o insuficiente de las personas que llegan a los conciertos en grupo.

2g. Muestra de conveniencia (no probabilística). Este esquema de muestreo no solo está sesgado por los caprichos del investigador, sino que también puede hacer que haya ausencia de respuesta por los asistentes que no puedan oír la invitación.

3a. La población en estudio (a la que los autores querían generalizar) era la población estadounidense de niños menores de 5 años en los años que estudiaron. Sabemos esto porque los autores utilizaron datos de una encuesta nacional para estimar la magnitud de la enfermedad por el metaneumovirus humano (MPVH) en Estados Unidos. Por supuesto, tendría mucho interés generalizar también a los años futuros, y muchos lectores lo harán sin meditar en ello. Sin embargo, es importante reconocer, especialmente con enfermedades infecciosas que pueden variar de unos años a otros, que la generalización más allá de los años del estudio es una inferencia adicional potencialmente frágil.

3b. La población accesible (la población de la que extrajeron los participantes) eran los niños < 5 años de edad que vivían en los condados que rodeaban a los tres centros de investigación (Cincinnati, Nashville y Rochester, NY) y que eran atendidos en dichos centros. Probablemente se seleccionaron estas ciudades por su proximidad a los investigadores. No está claro en qué medida son representativas de otras áreas de Estados Unidos en relación con la frecuencia de infección por el MPVH.

3c. El esquema de muestreo fue una muestra de conveniencia. La elección de los días de la semana (que no se especifican) podría haber llevado a algún sesgo si, por ejemplo, los padres de niños con síntomas respiratorios más leves durante el fin de semana esperan hasta el lunes para llevarlos a ver a un médico y si los síntomas del MPVH fueran más o menos intensos que los de otros virus. En los días en los que los investigadores estaban incluyendo pacientes, podían haber intentado obtener una muestra consecutiva (tampoco se especifica), lo que habría ayudado a controlar el sesgo de selección. No se muestra el motivo de la restricción de determinados meses del año, aunque probablemente fuera porque los autores pensaban que casi todos los casos de MPVH se producirían durante estos meses.

3d. Las observaciones estaban agrupadas por área geográfica, por lo que se debería tener en cuenta el agrupamiento por ciudad en el análisis estadístico. Cuanto más diferentes sean las estimaciones entre las ciudades, más amplios serán los intervalos de confianza. Intuitivamente, esto es lógico. Unas tasas por ciudad muy diferentes llevarían a preguntarse en qué medida habría sido diferente la estimación si se hubieran incluido otras ciudades, y esperaríamos ver reflejada esta incertidumbre en un intervalo de confianza más amplio.

Se produce un nivel más sutil de agrupamiento por año. Una vez más, si hay mucha variación de unos años a otros en la incidencia del MPVH, si se desea generalizar a los años futuros (en lugar de simplemente estimar cuál ha sido la incidencia en los años estudiados), también se debería incluir el agrupamiento por año en el análisis estadístico, y la variación significativa de la incidencia de unos años a otros también llevaría a un intervalo de confianza más amplio.

Capítulo 4 Planificación de las mediciones: precisión, exactitud y validez

1a. Dicotómica

1b. Continua

1c. Dicotómica

1d. Numérica discreta

1e. Nominal

1f. Numérica discreta

1g. Ordinal

1h. Continua

1i. Nominal

1j. Dicotómica

La potencia aumenta usando una variable de respuesta que contenga información ordenada. Por ejemplo, el último año de escolarización tiene más potencia que el grado universitario o más/menos que el grado universitario. De manera similar, el uso del índice de masa corporal como variable de respuesta continua ofrecería más potencia (contendría más información) para la mayoría de las preguntas de la investigación que la presencia o ausencia de obesidad. Una opción intermedia que se utiliza habitualmente es la variable ordinal normal/sobrepeso/obesidad.

2a. Este es un problema con la exactitud. Podría deberse a que un observador no visualiza correctamente la lectura (un segundo observador puede comprobar el resultado), pero es más probable que haya que ajustar la báscula.

2a. Es un problema con la precisión. La excesiva variabilidad podría ser un error del observador, pero es más probable que haya que ajustar la báscula.

2a. Esta situación puede reducir tanto la exactitud como la precisión. La exactitud se afectará porque, al sujetar el investigador al bebé, probablemente se alterará el peso observado; esto podría tender a aumentar constantemente el peso observado o a disminuirlo constantemente. Este problema con los participantes podría solventarse si la madre dedica un tiempo a calmar al niño; una alternativa sería pesar a la madre con y sin el bebé, y calcular la diferencia.

2a. Es fundamentalmente un problema de precisión, porque el indicador de la báscula variará alrededor del peso real (si la báscula es exacta). El problema se refiere a los participantes, y la solución es la misma que en el ejercicio 2c.

2a. El problema radica principalmente en la precisión, ya que los pesos de los bebés variarán, dependiendo de si comen o no y de si mojan los pañales antes de la medición. Este problema de variabilidad del participante puede reducirse advirtiendo a las madres que no alimenten a los niños durante 3 h antes de la exploración, y que los bebés estén desnudos.

2a. Validez predictiva: las puntuaciones de desgaste profesional predecían una variable de respuesta que podríamos esperar que se asociara al desgaste.

2a. Validez aparente: preguntar a las personas con qué frecuencia se sienten desgastadas parece un abordaje razonable para evaluar el síndrome de desgaste profesional.

2a. Validez de constructo: esta medida del desgaste es sensible a las circunstancias que cabría esperar que afectaran al desgaste profesional.

2a. Validez relacionada con el criterio: estas dos preguntas tienen un acuerdo estrecho con una medida estándar bien aceptada.

Capítulo 5 Preparación para el cálculo del tamaño de la muestra: hipótesis y principios subyacentes

1. Tamaño de la muestra = número proyectado de participantes en un estudio que el investigador necesita para poder detectar una determinada magnitud del efecto (como los niveles especificados de α y β).

Hipótesis nula = afirmación de la hipótesis de la investigación que indica que no hay diferencia entre los grupos que se comparan.

Hipótesis alternativa = afirmación de la hipótesis de la investigación que indica que hay una diferencia entre los grupos que se comparan.

Potencia = probabilidad de detectar una diferencia estadísticamente significativa entre los grupos que se comparan (con un tamaño de muestra determinado, con un nivel determinado de significación estadística) si la diferencia real en la población equivale a la magnitud del efecto.

Nivel de significación estadística = posibilidad preestablecida de rechazar falsamente la hipótesis nula.

Magnitud del efecto = mínimo tamaño de la diferencia en los dos grupos que se comparan que el investigador desea detectar.

Variabilidad = magnitud de la dispersión de una medición, que habitualmente se calcula como la desviación típica o el error típico de la media.

2a. Ninguno. Es un resultado estadísticamente significativo, y no hay nada que indique que representa un error de tipo I.

2b. El tamaño de la muestra era pequeño, y muy pocos participantes habrían presentado cáncer de pulmón durante el estudio. Estos resultados negativos se deben, casi con certeza, a un error de tipo II, especialmente dada la amplia evidencia en otros estudios de que el tabaquismo causa cáncer de pulmón.

2c. No hay razón epidemiológica ni fisiopatológica anterior para creer que el consumo de alcohol reduce el riesgo de sufrir diabetes; este resultado se debe, probablemente, a un error de tipo I. El investigador podría haber aportado más información: $p < 0,05$ podría ser $p = 0,04$ o $p = 0,001$; este último valor reduciría (aunque no descartaría) la probabilidad de error de tipo I.

Capítulo 6 Cálculo de la potencia y del tamaño de la muestra: aplicaciones y ejemplos

- H_0: no existe diferencia en el índice de masa corporal de los casos de cáncer de estómago y los testigos.

  H_A/bilateral): existen diferencias en el índice de masa corporal de los casos de cáncer de estómago y los testigos. El índice de masa corporal es una variable continua, y casos-testigos es dicotómica, por lo que se debe usar una prueba de la *t*.

  Magnitud el efecto = 1 kg/m^2

  Desviación típica = 2,5 kg/m^2

  $E/D = 0,4$

  Del apéndice 6A,

  Si $\alpha = 0,05$, $\beta = 0,20$, entonces se necesitan 100 participantes por grupo.

  Si $\alpha = 0,05$, $\beta = 0,10$, entonces se necesitan 133 participantes por grupo.

  Si $\alpha = 0,01$, $\beta = 0,20$, entonces se necesitan 148 participantes por grupo.

  Crédito extra: si el investigador solo tiene acceso a 60 casos, ¿cuál de las siguientes estrategias para aumentar la potencia será la más útil?

  - Usar una variable continua: el índice de masa corporal ya se está midiendo como una variable continua.
  - Usar una variable más precisa: tanto el peso como la altura son variables precisas, y la desviación típica del índice de masa corporal está compuesta principalmente por variación interindividual, que no se puede reducir. La normalización cuidadosa de las mediciones de la altura y el peso para reducir el error de medición seguiría siendo una buena idea, aunque no es la mejor opción.
  - Usar mediciones apareadas: no se puede aplicar; el «cambio» del índice de masa corporal no es relevante en esta situación.
  - Usar una variable de respuesta más habitual.
  - Usar tamaños de grupos desiguales: la *n* de los testigos puede aumentarse, ya que es fácil encontrar participantes sin cáncer de estómago. Por ejemplo, si el número de testigos puede aumentarse cuatro veces hasta 240, se puede usar la fórmula de aproximación de la página 69:

  $n' = ([c + 1]/2c) \times n$ donde n' representa el «nuevo» número de casos, *c* representa la proporción entre testigos y casos (en este ejemplo, 4) y *n* representa el «antiguo» número de casos (suponiendo un testigo por caso). En este ejemplo:

  $n' = ([4 + 1]/8) \times 100 = (5/8) \times 100 = 63$,

  que es aproximadamente el número de casos disponibles. Por lo tanto, un estudio con 60 casos y 240 testigos tendrá una potencia similar a la de uno con 100 casos y 100 testigos.

- H_0: no hay diferencia en la fuerza media entre los grupos tratados con DHEA y con placebo.

  H_A: existe una diferencia de la fuerza media entre los grupos tratados con DHEA y con placebo.

  $\alpha = 0{,}05$ (bilateral); $\beta = 0{,}10$

  Prueba = prueba de la *t*

  Magnitud del efecto = 10 % × 20 kg = 2 kg

  Desviación típica = 8 kg

  La magnitud del efecto normalizada *(E/D)* es 0,25 (2 kg/8 kg). Acuda al apéndice 6A, descienda por la columna izquierda hasta 0,25, luego recorra la tabla en horizontal hasta la quinta columna de la izquierda, donde α (bilateral) = 0,05 y $\beta = 0{,}10$. Se necesitarían aproximadamente 338 participantes por grupo. Si $\beta = 0{,}20$, el tamaño de la muestra es de 253 por grupo.

- H_0: no hay diferencia en el cambio medio de fuerza entre los grupos tratados con DHEA y con placebo.

  H_A: hay una diferencia en el cambio medio de la fuerza entre los grupos tratados con DHEA y con placebo.

  $\alpha = 0{,}05$ (bilateral); $\beta = 0{,}10$

  Prueba = prueba de la *t*

  Magnitud del efecto = 10 % × 20 kg = 2 kg

  Desviación típica = 2 kg

  La magnitud de la muestra normalizada *(E/D)* es de 1,0 (2 kg/2 kg). Acuda al apéndice 6A, descienda por la columna de la izquierda hasta 1,0, y luego vaya hasta la quinta columna desde la izquierda, donde α (bilateral) = 0,05 y $\beta = 0{,}10$. Se necesitarían, aproximadamente 23 participantes.

- H_0: no hay diferencia en cuanto a la frecuencia de ser zurdo entre los estudiantes disléxicos y los no disléxicos.

  H_A: existe una diferencia en cuanto a la frecuencia de ser zurdo entre los estudiantes disléxicos y los no disléxicos.

  $\alpha = 0{,}05$ (bilateral); $\beta = 0{,}20$

  Prueba = prueba de la χ^2 (ambas variables son dicotómicas)

  Magnitud del efecto = razón de posibilidades de 2,0

  Dado que la proporción de estudiantes no disléxicos que son zurdos (P_2) es de alrededor de 0,1, el investigador desea poder detectar una proporción de estudiantes disléxico que sean zurdos (P_1) que proporcione una razón de posibilidades de 2,0. El cálculo del tamaño de la muestra usará una prueba de la χ^2, y es preciso usar el apéndice 6B. Sin embargo, ese apéndice está pensado para introducir las dos proporciones, no la razón de posibilidades, y todo lo que se conoce es una de las proporciones ($P_2 = 0{,}1$).

  Para calcular el valor de P_1 que da una razón de posibilidades de 2, se puede usar la fórmula de la página 59:

  $P_1 = \text{RP} \times P_2/([1 - P_2] + [\text{RP} \times P_2])$

  En este ejemplo:

  $P_1 = (2 \times 0{,}1)/([1 - 0{,}1] + [2 \times 0{,}1]) = 0{,}18$

  Por lo tanto, $P_1 = 0{,}18$ y $P_2 = 0{,}1$. $P_1 - P_2 = 0{,}08$.

  La tabla 6B-2 del apéndice 6B señala un tamaño de la muestra de 318 por grupos.

Crédito extra: Inténtelo usando la fórmula de la página 78; trabaje solo sacando seis decimales. Luego, obtenga una respuesta instantánea de la calculadora de nuestra página web: www.epibiostat.ucsf.edu/dcr/.

- La desviación típica de las puntuaciones de CI es alrededor de la cuarta parte del valor «habitual» (que es 170 - 130 = 40 puntos), o 10 puntos.

  La amplitud total del intervalo de confianza = amplitud total/desviación típica,

  A/D = 0,6

  Usando la tabla 6D, descienda por la columna de *A/D* hasta 0,60, y luego hasta el intervalo de confianza del 99 %. Se necesitarían en promedio alrededor de 74 puntuaciones de CI de estudiantes de medicina para obtener una puntuación media con el intervalo de confianza especificado.

Capítulo 7 Diseño de estudios transversales y de cohortes

1a. Medir las concentraciones séricas de vitamina B_{12} en una cohorte de personas de más de 70 años de edad y sin antecedente de fractura de cadera, seguirlas durante un período de tiempo (p. ej., 5 años) para detectar la aparición de fracturas de cadera, y después analizar la asociación entre las concentraciones de vitamina B_{12} y las fracturas de cadera incidentes. (Se podría realizar un estudio de menor tamaño, aunque menos generalizable, estudiando solo a mujeres, que tienen mayor incidencia de fracturas de cadera; un estudio de incluso menor tamaño permitiría incluir solo a mujeres blancas, que tienen la máxima incidencia de fracturas.)

1b. Una ventaja del diseño de cohortes prospectivo para el estudio de la carencia de vitamina B_{12} y las fracturas de cadera:

- La secuencia temporal (la fractura de cadera sigue a la carencia de vitamina B_{12}) ayuda a establecer una relación de causa-efecto. Las personas que se fracturan la cadera podrían tener carencia de vitamina B_{12} después de la fractura por reducir el consumo de vitamina B_{12}, tal vez por su ingreso en una residencia de ancianos.

Un inconveniente del diseño de cohortes prospectivo:

- Un estudio de cohortes prospectivo necesitará muchos participantes y un seguimiento de muchos años. El estudio será, por lo tanto, caro, y los hallazgos son tardíos.

1c. Podría realizarse un estudio de cohortes retrospectivo si pudiera encontrar una cohorte con suero y con un seguimiento razonablemente completo para determinar quién sufrió fractura de cadera. La principal ventaja de este diseño es que consumiría menos tiempo y dinero. El principal fallo es que la medición de la vitamina B_{12} podría alterarse por su almacenamiento a largo plazo, y que puede no disponerse de las determinaciones de posibles factores de confusión (actividad física, tabaquismo, etc.).

2a. Aunque el estudio PRIDE es un ensayo aleatorizado, la descripción de la exploración inicial es un estudio transversal (observacional). Los estudios transversales muchas veces son el primer paso en los estudios de cohortes y los ensayos aleatorizados.

2b. Aunque es posible que la depresión aumente la incontinencia urinaria, también parece, al menos, igual de aceptable que la incontinencia urinaria aumente el riesgo de depresión. Como se analizará en el capítulo 9, también es posible que la asociación se deba a sesgo, es decir, si las mujeres deprimidas tuvieran más probabilidad de referir episodios de incontinencia, incluso aunque ya no los tuvieran, o a confusión, si un tercer factor (p. ej., gravedad de la obesidad) produjera tanto depresión como incontinencia.

Un estudio longitudinal (de cohortes) podría ayudar a clarificar la secuencia temporal de la asociación. Por ejemplo, se podría seguir a mujeres deprimidas y no deprimidas con poca incontinencia o sin incontinencia en situación inicial, para ver si las mujeres deprimidas presentan más incontinencia, o una incontinencia peor, a lo largo del tiempo. De manera similar, se podría seguir a mujeres continentes e incontinentes sin antecedentes de depresión para determinar si las mujeres con más incontinencia tienen más probabilidad de deprimirse. Finalmente, y lo que es más convincente, los investigadores podrían estudiar los cambios de la depresión o la incontinencia, ya se produzcan de manera natural o (idealmente) como consecuencia de una intervención, y ver si los cambios de una preceden a los cambios de la otra. Por ejemplo, ¿los síntomas depresivos mejoran cuando se trata con éxito la incontinencia? ¿La incontinencia (referida) mejora cuando se alivia la depresión?

Capítulo 8 Diseño de estudios de casos y testigos

1a. Los casos podrían ser todas las mujeres entre 30 y 75 años de edad con cáncer de ovario comunicado a un registro local, y con las que se puede establecer contacto telefónico y que pueden dar su acuerdo para participar.

1b. Las testigos podrían ser una muestra aleatoria de todas las mujeres de entre 30 y 75 años de edad de los mismos condados que en el registro de tumores. La muestra aleatoria podría obtenerse mediante una llamada telefónica aleatoria (de aquí la necesidad de restringir los casos a las que tienen teléfono).

1c. Dado que el cáncer de ovario necesita un tratamiento intensivo y puede ser mortal, algunos casos puedan ser reacios a participar en el estudio o pueden haber muerto antes de poder entrevistarles. Si un antecedente familiar de cáncer de ovario está relacionado con formas más agresivas de este tipo de cáncer, el estudio podría infravalorar su riego relativo, porque esos casos con un antecedente familiar positivo tendrían menos probabilidad de sobrevivir lo suficiente como para incluirlas en la muestra de casos.

Si un cáncer de ovario familiar es más benigno que otros cánceres de ovario, podría ocurrir lo contrario. Igualmente, es posible que mujeres sanas con una familiar con cáncer de ovario estuvieran más interesadas en el estudio y más dispuestas a participar como testigos. En esa situación, la prevalencia del antecedente familiar de cáncer de ovario en el grupo testigo sería artificialmente elevada, y el cálculo del riesgo de sufrir cáncer de ovario por antecedentes familiares sería falsamente bajo. Este problema podría reducirse al mínimo si no se comunica a las posibles participantes como testigos cuál es exactamente la pregunta de la investigación, o qué cáncer exactamente es el que se va a estudiar, si puede realizarse de un modo que sea aceptable para el comité de participantes humanos.

1d. El antecedente familiar de cáncer de ovario suele medirse preguntando a las participantes cuántos familiares femeninos tienen y cuántas han sufrido cáncer de ovario. Recuerde que el sesgo puede ser un problema con este método. Las mujeres con cáncer de ovario, que puedan preocuparse por la posibilidad de una predisposición genética a su enfermedad, pueden recordar con más probabilidad a las familiares con cáncer de ovario que las mujeres sanas que no tienen motivos para pensar en esa posibilidad. Esto haría que la estimación de la asociación entre el antecedente familiar y el cáncer de ovario fuera falsamente elevada.

Además, las mujeres pueden confundir los cánceres ginecológicos (cervical, uterino y ovárico), y confundir también los tumores ginecológicos benignos que requieren una intervención quirúrgica con los tumores malignos. Esto puede llevar a una clasificación errónea (algunas mujeres sin antecedentes familiares de cáncer de ovario comunicarán tener el factor de riesgo, y se las clasificará erróneamente). Si esta clasificación errónea se produce tanto en los casos como en los testigos, el cálculo de la asociación entre antecedentes familiares y cáncer de ovario será falsamente bajo. Si este tipo de clasificación errónea es más frecuente en los casos (en los que hay más probabilidad de malinterpretar el tipo de cáncer o el motivo de la cirugía en las familiares), la estimación de la asociación entre antecedentes familiares y cáncer de ovario será falsamente elevada. La clasificación errónea puede disminuirse comprobando los registros anatomopatológicos de las familiares que se ha comunicado que han tenido cáncer de ovario, con el fin de verificar el diagnóstico.

Finalmente, sería deseable tener en consideración la *oportunidad* de los casos y los testigos de tener un antecedente familiar positivo: las mujeres con muchas hermanas mayores tienen más oportunidad de tener antecedentes familiares positivos que las que solo tienen hermanos o hermanas menores. Como se señala en el capítulo 9, el emparejamiento y la estratificación son dos formas de abordar esta posibilidad.

1e. El método más sencillo sería dicotomizar el antecedente familiar de cáncer de ovario (p. ej., familiares en primer grado o no) y usar la razón de posibilidades como medida de asociación. La razón de posibilidades se aproxima al riesgo relativo, porque la variable de respuesta (cáncer de ovario) no es frecuente. Una prueba de la χ^2 sería una prueba adecuada de significación estadística. Por otro lado, si se cuantificaron los antecedentes familiares (proporción de familiares de primer y segundo grado afectadas), se podría buscar una relación dosis-respuesta, calculando razones de posibilidades para cada nivel de exposición.

1f. El diseño de casos y testigos es un modo razonable de responder a esta pregunta de la investigación, a pesar de los problemas del sesgo de muestreo, sesgo de recuerdo y clasificación errónea que se señalaron anteriormente. La principal alternativa sería un estudio de cohortes extenso; sin embargo, como el cáncer de ovario es muy infrecuente, probablemente no sea viable un diseño de cohortes para responder esta pregunta específica. Sería ideal un estudio de cohortes retrospectivo en el que ya estuvieran recogidos sistemáticamente los datos de los antecedentes familiares, si se pudiera encontrar dicha cohorte.

2a. Los casos podrían ser conductores jóvenes (tal vez de 16 a 20 años de edad) que hubieran tenido accidentes de tráfico, y los testigos podrían ser amigos o conocidos a los que identifiquen. Sería importante excluir a amigos con los que hayan jugado a videojuegos para evitar un emparejamiento excesivo. Es probable que las llamadas telefónicas con marcado aleatorio fueran menos eficaces como estrategia para identificar a los testigos, debido a la elevada prevalencia de teléfonos móviles (que, al contrario de las líneas terrestres, no están localizados en un ámbito geográfico determinado) en este grupo de edad. También se podrían identificar los casos y los testigos si el investigador tuviera acceso a los registros de una compañía de seguros de automóvil. Se podría proponer que los casos y los testigos estuvieran emparejados por sexo, debido a que tanto jugar con videojuegos como tener accidentes de coche son más frecuentes en hombres jóvenes. La exposición se podría medir con un cuestionario o una entrevista sobre el uso de videojuegos. Sería importante preguntar por videojuegos que no supongan conducción, además de aquellos que sí lo hacen, porque la inferencia causal mejoraría si la asociación fuera específica, es decir, si hubiera un efecto para el uso de juegos de conducción o carreras, pero no para juegos de

disparos o de otro tipo.

2b. Para exposiciones intermitentes que se supone que tienen un efecto breve, como el uso de un videojuego inmediatamente antes de conducir, los estudios cruzados son una opción atractiva. Como en el ejercicio 2a, los casos podrían ser conductores jóvenes que hubieran tenido accidentes de tráfico. En los estudios cruzados no hay testigos, simplemente períodos temporales testigo. Así, se preguntaría a los conductores que son casos por el uso de videojuegos inmediatamente antes del viaje en el que se produjo el accidente, y también por períodos temporales de control en los que no tendrán accidentes. Se compara el período temporal inmediatamente anterior al accidente en un análisis emparejado con otros períodos temporales para ver si el uso de videojuegos de carreras era más frecuente en el período previo al accidente que en otros períodos temporales.

Capítulo 9 Mejora de la inferencia causal en estudios de observación

- Hay cinco posibles explicaciones para la asociación entre el consumo de frutas y verduras con la dieta y la CI:
    - Casualidad: el hallazgo de que las personas con CI consumen menos frutas y verduras se debió a un error aleatorio. Como se comenta en el capítulo 5, el valor de *p* permite cuantificar la magnitud de la diferencia observada en comparación con la que podría haberse esperado solo por casualidad; el intervalo de confianza del 95 % muestra el intervalo de valores compatibles con los resultados del estudio. Siendo todo lo demás igual, cuanto menor sea el valor de *p* y más lejos esté el valor nulo del extremo más próximo del intervalo de confianza, menos posible es la explicación de la casualidad.
    - Sesgo: había un error sistemático (una diferencia entre la pregunta de la investigación y el modo en que se realizó el plan del estudio) con respecto a la muestra, la variable predictiva o la variable de respuesta. Por ejemplo, puede haber sesgo en la muestra si los testigos eran pacientes del mismo plan sanitario que los casos, pero se escogieron entre los que acudieron a una exploración anual de mantenimiento de la salud, ya que estos pacientes pueden ser más conscientes de su salud (y, por tanto, comer más frutas y verduras) que toda la población con riesgo de sufrir CI. Las mediciones de la dieta pueden estar sesgadas si las personas que han sufrido un infarto recuerdan mejor los hábitos dietéticos que los testigos (sesgo de recuerdo), o si entrevistadores que conocen la asignación realizan las preguntas o registran las respuestas de forma diferente en los casos y en los testigos.
    - Efecto-causa: es posible que, tras sufrir un infarto, cambiaran las preferencias dietéticas de las personas, de manera que consumieran menos frutas y verduras que antes del infarto de miocardio. La posibilidad de efecto-causa puede analizarse a menudo diseñando variables para examinar la secuencia histórica (p. ej., preguntando a los casos y los testigos por su dieta previa, en lugar de su dieta actual).
    - Confusión: puede haber otras diferencias entre los que comen más frutas y verduras y los que comen menos, y estas otras diferencias pueden ser la causa real de la menor incidencia de CI. Por ejemplo, las personas que comen más frutas y verduras pueden hacer más ejercicio.

      En la tabla siguiente se resumen los posibles abordajes para controlar la confusión por el ejercicio:

Método	Plan	Ventajas	Inconvenientes
Fase de diseño			
Especificación	Incluir solo personas que comunican que no hacen ejercicio habitual	Sencillo	Limitará el número de participantes elegibles, dificultando la inclusión. El estudio no puede generalizarse a los que hacen ejercicio
	Emparejar cada caso a un testigo	Elimina el efecto del ejercicio como factor predictivo de CI, a menudo con un ligero	Precisa un esfuerzo extra para identificar testigos que emparejar a cada caso. Desperdiciará casos si no hay un

Emparejamiento	con un nivel de ejercicio similar	aumento de precisión (potencia) para observar la dieta como variable predictiva	testigo con nivel de ejercicio similar. Elimina la oportunidad de estudiar el efecto del ejercicio sobre la CI
Fase de análisis			
Estratificación	Para el análisis, agrupar los participantes en tres o cuatro estratos de ejercicio	Fácil, comprensible y reversible	Solo puede evaluar razonablemente algunos estratos y algunas variables de confusión. Perderá parte de la información contenida en el ejercicio, medido como una variable continua cambiando a una variable categórica, y esto puede causar un control incompleto de la confusión
Ajuste estadístico (modelado)	Usar un modelo de regresión logística para controlar la forma física y otras posibles variables de confusión	Puede controlar de forma reversible toda la información sobre la forma física como una variable predictiva continua, mientras controla simultáneamente otras posibles variables de confusión, como la edad, la raza y el tabaquismo	El modelo estadístico podría no encajar en los datos, causando un control incompleto de la confusión y unos resultados posiblemente engañosos. Por ejemplo, el efecto de la dieta o la forma física puede no ser igual en fumadores y en no fumadores. Las posibles variables de confusión importantes deben haberse medido por adelantado. A veces es difícil entender y describir los resultados del modelo, especialmente cuando las variables no son dicotómicas

Además de estas cuatro estrategias para controlar la confusión en los estudios de observación, hay una solución final: diseñar un ensayo clínico aleatorizado con enmascaramiento.

- Causa-efecto: la quinta explicación posible es que comer frutas y verduras realmente reduce el índice de episodios coronarios. A esta explicación se llega probablemente, en parte, por un proceso de exclusión, alcanzando el juicio de que cada una de las otras cuatro explicaciones no es probable y, en parte, buscando otras pruebas para apoyar la hipótesis causal. Ejemplos de esto último son los datos biológicos de que hay componentes de las frutas y las verduras (p. ej., antioxidantes) que protegen frente a la ateroesclerosis, y estudios ecológicos en los que se ha visto que la CI es mucho menos frecuente en poblaciones que consumen más frutas y verduras.

- Este es un ejemplo de condicionamiento a un efecto compartido: en el estudio se incluyó solo a lactantes con fiebre, que puede estar producida por infecciones urinarias e infecciones óticas. Como los niños no circuncidados tenían una probabilidad mucho mayor de tener una infección urinaria, tenían más probabilidad de tener una causa de la fiebre distinta a una infección ótica (es decir, estaban sobrerrepresentados en los niños que no tenían una infección de oído).

- La asociación entre el uso de paracetamol por la madre y el asma en la descendencia se podría evaluar en un estudio de cohortes en el que se preguntara a las madres por el consumo de paracetamol durante el embarazo y se siguiera a la descendencia para detectar la aparición de asma. Los investigadores buscarían datos de que el genotipo materno modifica el efecto de la exposición de la madre al paracetamol sobre el asma en los niños (interacción), con una asociación mayor entre la exposición y la variable de respuesta en aquellas que se predijera que tenían más susceptibilidad genética. De hecho, esto es lo que publicaron Shaheen y cols. (*J Allergy ClinImmunol* 2010;126(6):1 141-1 148 e7.) en el estudio *Avon Longitudinal Study of Parents and Children* (ALSPAC).

Capítulo 10 Diseño de un estudio aleatorizado con enmascaramiento

1a. La principal ventaja de usar el biomarcador (una variable continua) como variable de respuesta principal del ensayo es un menor tamaño de la muestra y una duración más corta para determinar si el tratamiento reduce la concentración del marcador. El inconveniente

es la duda de si el cambio de la concentración del marcador inducido por el tratamiento significa que el tratamiento reducirá la incidencia de progresión de la variable de respuesta con más importancia clínica, la aparición de demencia.

1b. El diagnóstico clínico de demencia es una respuesta más significativa del ensayo que podría mejorar la práctica clínica para la prevención de la demencia. La desventaja es que un ensayo de este tipo sería grande, prolongado y costoso.

2a. Debe preguntarse a los participantes en todas las consultas de seguimiento si han sufrido diarrea, náuseas o vómitos. Podría hacerse usando un formato de casillas de verificación, que es fácil de codificar y de analizar. Para encontrar otros efectos adversos no previstos, debe pedirse también a los participantes que describan otros síntomas, afecciones o asistencia médica (como una hospitalización o nuevos fármacos recetados) que se hayan producido desde la consulta anterior. Estas preguntas se plantearían mediante un método abierto, y las respuestas posteriormente se clasificarían y categorizarían para el análisis de los datos.

2b. La recopilación de datos basales debe incluir: *1)* información sobre cómo contactar con el participante, un amigo íntimo, un familiar o el médico, para un mejor seguimiento; *2)* características de la población incluida (como edad, etnicidad/raza y sexo) para poder describir la cohorte del estudio; *3)* factores de riesgo de la respuesta (como hipertensión o antecedentes familiares de demencia) que pudieran identificar a participantes con la mayor tasa de respuesta y que se podrían utilizar para demostrar que los grupos en estudio eran comparables en situación inicial, además de definir subgrupos para los análisis secundarios, y *4)* medición de la variable de respuesta (gravedad del deterioro cognitivo). Deben almacenarse muestras biológicas para permitir mediciones futuras de factores, como genotipos de enzimas que metabolizan el fármaco, que podrían influir en la eficacia del tratamiento.

La aleatorización estratificada por bloques podría garantizar que hubiera un número similar de participantes con el genotipo Apoε4 en el grupo de tratamiento y en el grupo con placebo. Esto podría ser especialmente importante si el efecto del tratamiento se ve influido por la presencia del genotipo. Por otro lado, este proceso complica el ensayo (valorando el genotipo Apoε4 antes de la participación, se retrasará la aleatorización, y surgen problemas como la mejor forma de aconsejar a los participantes sobre los resultados). El riesgo de un desequilibrio importante en un ensayo relativamente grande (> 200 por grupo) es bajo, de modo que la aleatorización simple sería una buena elección.

Capítulo 11 Otros diseños de ensayos clínicos y problemas en su realización

1. El principal objetivo de un ensayo de fase I es determinar si el tratamiento es suficientemente seguro y bien tolerado como para que se puedan hacer otros ensayos a fin de encontrar la mejor dosis y evaluar su eficacia clínica. En un ensayo de fase I se incluiría a hombres con calvicie de patrón masculino, y se usaría una o más posibles dosis de tratamiento (aumentando la dosis solo si la dosis previa no produjo efectos adversos) con la principal respuesta de efectos adversos, como la aparición de exantema. No habría grupo testigo.

2a. La utilidad de la comparación de finasterida con HairStat depende, en gran medida, de la solidez que tengan los datos que respaldan el uso de finasterida, tratamiento estándar de la calvicie con patrón masculino. Si estos datos no son muy sólidos, o si la finasterida no se utiliza habitualmente en la práctica clínica, sería mejor comparar HairStat con un placebo. Un ensayo aleatorizado con un placebo permitirá tener datos claros de que HairStat es mejor que el placebo. Puede ser razonable comparar HairStat con finasterida si se considera que la finasterida es el tratamiento estándar de la calvicie con patrón masculino y si hay ensayos aleatorizados de buena calidad que documentan la eficacia de la finasterida. En este caso, los investigadores deben decidir primero si piensan que HairStat es más eficaz que la finasterida. De ser así, un estudio con comparador activo sería la mejor opción para comparar HairStat con finasterida. Si los investigadores piensan que HairStat es igual de eficaz que la finasterida, aunque será mucho más económico, deben plantearse un estudio de ausencia de inferioridad. En este caso, deben tener cuidado de utilizar un diseño del estudio que sea muy similar al que se utilizó para documentar la eficacia de la finasterida (criterios de inclusión, dosis, duración del tratamiento, medidas de la respuesta), y deben realizar el estudio para garantizar que haya un incumplimiento y una pérdida durante el seguimiento mínimos. Un inconveniente importante de los estudios de ausencia de inferioridad es que es probable que el tamaño de la muestra sea mucho mayor de lo que es necesario para un ensayo controlado con placebo.

2b. Un diseño factorial que incluye un placebo tiene las ventajas de comparar cada tratamiento con el placebo, y (si se planifica con una potencia estadística adecuada) comprobar si la combinación de los tratamientos es mejor que cada uno por separado. Los inconvenientes son el mayor tamaño, el mayor coste y la complejidad del ensayo.

3. El cumplimiento con las consultas, el protocolo y los medicamentos del estudio puede mejorarse:

- Empleando personal de la investigación agradable que se muestre entusiasta con el estudio
- Recordando (por mensajes digitales, por teléfono, correo postal o correo electrónico) las consultas y la importancia de cumplir el tratamiento
- Reembolsando los gastos de trasporte, aparcamiento y otros relacionados con el estudio

- Dos visitas de selección antes de la aleatorización para identificar a los pacientes que tienen más probabilidad de no acudir a las visitas de seguimiento
- Un período de preinclusión en el que se pide a los participantes que utilicen el gel capilar de placebo, y en el que se excluye a los que no cumplan el tratamiento
- Otras posibles estrategias enumeradas en la tabla 11-2

4. El principal inconveniente del análisis de intención de tratar es que incluye participantes que no han cumplido el tratamiento aleatorizado, y que, por lo tanto, reducen la magnitud aparente de cualquier efecto que se observa en todo el grupo aleatorizado. Sin embargo, los inconvenientes de usar el análisis de «como se les trató» en lugar del de intención de tratar son incluso mayores. Porque los participantes que no cumplen la intervención suelen ser distintos de los que cumplen en aspectos importantes, pero no medidos, el análisis de «como se les trató» ya no tiene una comparación aleatorizada verdadera y se puede llegar a la conclusión incorrecta de que HairStat es eficaz.

5. La conclusión de que HairStat funciona mejor en varones jóvenes, según un análisis de subgrupos, puede ser errónea y engañosa, porque el resultado puede deberse al azar. La probabilidad de hallar un efecto «significativo» en un subgrupo cuando no hay ningún efecto general significativo aumenta con el número de grupos estudiados; no está claro cuántos subgrupos se estudiaron para hallar este efecto «significativo». La afirmación de que el tratamiento es eficaz en varones de menos de 40 años implica que el tratamiento no es eficaz (o incluso tiene el efecto contrario) en varones de más edad. Este resultado también debe comunicarse y verificarse estadísticamente para detectar una modificación del efecto sobre el crecimiento del cabello con HairStat debido a la edad. La afirmación de que HairStat es eficaz en el subgrupo de hombres más jóvenes solo se debe hacer si los análisis de subgrupos se hubieran especificado por adelantado (en situación ideal, basándose en un fundamento biológico para sospechar que HairStat podría ser más eficaz en hombres jóvenes), si no se hubiera estudiado un número grande de subgrupos, y si el valor de *p* para la modificación del efecto (interacción) entre el efecto del tratamiento y la edad fuera estadísticamente significativo.

Capítulo 12 Diseño de estudios de pruebas médicas

1a. La mejor forma de obtener una muestra de participantes para estudiar una prueba diagnóstica suele ser muestrear pacientes con riesgo de sufrir una enfermedad, antes de que se sepa quién tiene la enfermedad y quién no. En este caso, lo mejor sería probablemente muestrear mujeres que acuden a una consulta o a un servicio de urgencias con dolor abdominal compatible con enfermedad inflamatoria pélvica (EIP). Comparar la velocidad de sedimentación globular (VSG) de las mujeres hospitalizadas con EIP con la de una población testigo sana sería el peor método, porque tanto el espectro de la enfermedad como, y especialmente, el espectro de la ausencia de enfermedad no son representativos de los grupos en los que se usaría clínicamente la prueba. (Las hospitalizadas por EIP tienen probablemente una enfermedad más grave que el promedio, y las voluntarias sanas tienen mucha menos probabilidad de que los valores de su VSG estén elevados que las mujeres con dolor abdominal por otras causas distintas a la EIP.)

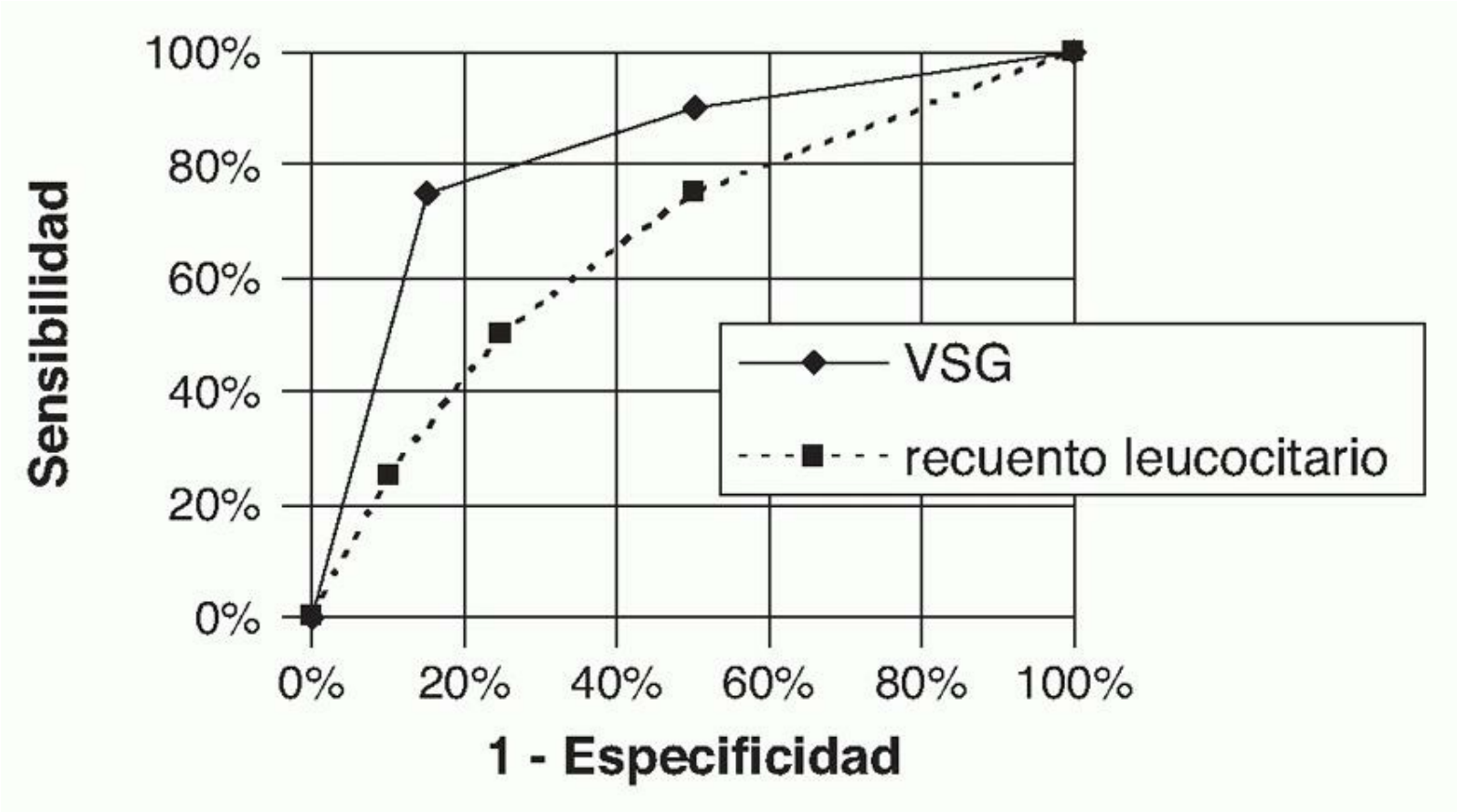
1b. Si las personas que asignaban el diagnóstico final usaron la VSG para decidir quién tenía EIP y quién no, tanto la sensibilidad como la especificidad podrían estar falsamente elevadas. Cuanto más se base la asignación del diagnóstico en la VSG, mayor será el sesgo (denominado «sesgo de incorporación») del estudio.

1c. La mejor respuesta es que no debe usar ningún valor corte particular para definir un resultado anormal. En su lugar, debe representar gráficamente la interrelación entre la sensibilidad y la especificidad usando una curva de eficacia diagnóstica (curva ROC), y presentar razones de verosimilitudes para varios intervalos de VSG (p. ej., < 20 mm/h, 20 a 49 mm/h, ≥ 50 mm/h), en lugar de los valores de sensibilidad y especificidad en diferentes puntos de corte. Se ilustra en la tabla siguiente, que puede crearse a partir de la información de la pregunta:

VSG	EIP	No EIP	Razón de verosimilitudes
≥ 50	75%	15%	5,00
20-49	15%	35%	0,43
< 20	10%	50%	0,20

100% 100%

La curva ROC también podría usarse para comparar la VSG con otras pruebas, como el recuento leucocitario. Se ilustra en la hipotética curva que aparece a continuación, que sugiere que la VSG es superior al recuento leucocitario para predecir la EIP:



Herramientas de imágenes

2a. Este problema ilustra el habitual error de excluir personas del numerador sin excluirlas del denominador. Aunque es cierto que solo había dos niños con lesiones intracraneales «inesperadas», el denominador para el rendimiento debe ser el número de niños en los que cabría esperar que las lesiones intracraneales fueran inesperadas, es decir, aquellos con exploración neurológica y estado mental normales. Esto probablemente sea un número mucho menor de 200. Por ejemplo, suponga que solamente el 50 % de los niños en los que se solicitó una TC tenían un estado mental normal y no tenían hallazgos neurológicos. En esa situación, el rendimiento sería 2 a 50, o el 4 %, es decir, casi cuatro veces mayor.

2b. Salvo que el hallazgo de una lesión intracraneal cause cambios en el tratamiento y exista algún modo de calcular los efectos de esos cambios del tratamiento sobre la respuesta, será muy difícil saber qué rendimiento es suficiente para que esté justificada la TC. Sería mejor usar «lesión intracraneal que precisa una intervención quirúrgica» como variable de respuesta de este estudio, aunque se necesitará un acuerdo sobre qué lesiones necesitan ser intervenidas y un cálculo de la eficacia de estas intervenciones para mejorar la respuesta.

2c. La principal ventaja de estudiar los efectos de la TC sobre las decisiones clínicas es la capacidad de examinar los posibles efectos beneficiosos de los resultados normales. Por ejemplo, una TC normal podría cambiar el plan de tratamiento de «ingreso para observación» a «alta». En los estudios de rendimiento diagnóstico suele suponerse que los resultados normales tienen poco valor. En segundo lugar, como se comentó anteriormente, los resultados anormales de la TC podrían no llevar a ningún cambio en el tratamiento (p. ej., si no se precisaba neurocirugía y el paciente iba a ingresar de todos modos). Estudiar los efectos de las pruebas sobre la toma de decisiones médicas ayuda a determinar la cantidad de información útil que proporcionan, más allá de la que ya se conoce cuando se solicita la prueba.

3a. Si solo se incluyen niños a los que se realizó una TC, el estudio podrá sufrir sesgo de verificación (apéndice 12B), en el que la sensibilidad está falsamente aumentada y la especificidad está falsamente disminuida, porque los niños sin anomalías neurológicas focales (que son «negativos falsos» o bien «negativos verdaderos») estarán poco representados en este estudio.

3b. Si se incluyen los niños con lesiones craneales a los que no se realizó una TC, y se supone que no tienen lesión intracraneal si se recuperan sin neurocirugía, el estudio podrá sufrir sesgo de verificación diferencial («sesgo de método de referencia doble»; apéndice 12C), que tenderá a aumentar tanto la sensibilidad como la especificidad, si algunas lesiones intracraneales se resuelven sin neurocirugía.

Capítulo 13 Investigación utilizando datos existentes

- Algunas posibilidades:
    - Analizar datos del estudio *National Health and Nutrition Examination Survey* (NHANES). Estos estudios poblacionales nacionales se realizan periódicamente, y cualquier investigador puede disponer de sus resultados con un coste nominal. Contienen datos de muestras que incluyen variables sobre la historia clínica de colecistopatía comunicada por el propio paciente y los resultados de la ecografía abdominal.
    - Analizar datos de Medicare sobre la frecuencia de colecistectomía en pacientes de más de 65 años de edad en Estados Unidos, o datos de la encuesta *National Hospital Discharge Survey* sobre la frecuencia de esa intervención a todas las edades. Ambas series de datos contienen una variable para la raza. Los denominadores deben proceder de datos del censo. Al igual que en el estudio NHANES, son muestras muy buenas basadas en la población, pero tienen el problema de que responden a una pregunta de la investigación algo diferente (cuál es la incidencia de tratamiento quirúrgico por una colecistopatía). Esto puede ser diferente de la incidencia real de colecistopatía, debido a factores como el acceso a la asistencia.

  2a. Las principales ventajas son que usar datos del estudio CHS en un análisis de datos secundario era rápido, fácil y barato, especialmente en comparación con el tiempo y el coste de planificar y realizar un gran estudio de cohortes. Además, el compañero investigador ha desarrollado una colaboración con los investigadores del estudio CHS y ha podido añadir medidas más sofisticadas de función renal al estudio CHS en un estudio auxiliar.

  2b. En algunos casos, la serie de datos secundarios no proporciona mediciones óptimas de las variables predictiva y de respuesta, ni de las posibles variables de confusión. Es importante estar seguro de que la serie de datos proporcionará respuestas razonables a la pregunta de investigación antes de invertir el tiempo y el esfuerzo necesarios para lograr el acceso a los datos. Otro inconveniente es que puede ser difícil obtener datos de algunos estudios; el investigador generalmente tiene que escribir una propuesta, encontrar un colaborador que sea coinvestigador del estudio, y obtener la autorización del comité directivo del estudio y del promotor.

- Se han realizado varios grandes ensayos aleatorizados y controlados sobre el efecto de los estrógenos y los moduladores selectivos del receptor estrogénico en varias respuestas de enfermedad, como cáncer, episodios cardiovasculares y episodios tromboembólicos. Estos ensayos son los ensayos aleatorizados *Women's Health Initiative, Breast Cancer Prevention Trial, Multiple Outcomes of Raloxifene Evaluation* y *Raloxifene Use for the Heart*. La mejor forma de empezar para este investigador sería determinar si puede medir los estrógenos en el suero congelado almacenado y, si es así, determinar si alguno de estos grandes ensayos tiene suero almacenado que pudiera usarse para esta medición. El mejor diseño para esta pregunta es un estudio de casos y cohortes o de casos y testigos anidado. El investigador necesitará, probablemente, redactar una propuesta para este estudio auxiliar, conseguir la aprobación del comité directivo del ensayo y del patrocinador, y lograr la financiación para realizar las mediciones; un proyecto relativamente barato, ya que la mayor parte del coste del estudio ya se realizó en el ensayo principal.

Capítulo 14 Abordaje de cuestiones éticas

1a. Depende de si los participantes del estudio original dieron su consentimiento para que se utilizaran sus muestras para el secuenciado del ADN, de si dieron su consentimiento para que las mediciones del ADN se utilizaran en estudios futuros, y qué tipos de estudios futuros se especificaron. El consentimiento original no incluiría la investigación propuesta si las muestras de sangre se hubieran obtenido para utilizarlas únicamente para repetir las pruebas especificadas en el protocolo en caso de que se perdieran las muestras o hubiera accidentes de laboratorio (como el colesterol y la hemoglobina A1c). De manera similar, el consentimiento original no incluiría la investigación propuesta si los participantes hubieran dado su consentimiento para que se utilizaran las muestras de sangre para la medición genética del ADN en estudios futuros de arteriopatía coronaria, pero no hubiera ninguna mención a la utilización de las muestras en estudios de diabetes.

1b. Según la legislación federal, un estudio puede realizarse con datos y muestras existentes si el nuevo investigador no puede identificar a los participantes, bien directamente o bien con la ayuda de alguien más. Así, si el nuevo investigador recibe muestras y datos etiquetados solo por el número de ID, y el código que vincula las muestras y las identidades de los participantes se ha destruido o no está a disposición del nuevo investigador, no es necesario obtener un consentimiento informado para el estudio secundario. La justificación

ética es que hacer que el material sea anónimo de esta manera protege a los participantes de las violaciones de la confidencialidad. La suposición es que nadie podría oponerse a que se use su material o sus datos si no hay ningún riesgo de violación de la confidencialidad. Obsérvese, no obstante, que algunos participantes podrían poner objeciones a que alguien quisiera secuenciar su ADN, incluso manteniendo la confidencialidad, ya que el ADN contiene información que podría llevar, finalmente, a la pérdida de la confidencialidad.

1c. Cuando los investigadores recopilan nuevas muestras en un proyecto de investigación, es prudente pedir autorización para recoger y guardar más sangre para usarla en futuros estudios de investigación. Guardar muestras permite realizar futuras investigaciones de un modo más fiable que reuniendo una nueva cohorte. Se recomienda el consentimiento gradual: se pide el consentimiento al participante: *1)* para el estudio específico (p. ej., el estudio de cohortes original); *2)* para otros proyectos de investigación del mismo tema general (como riesgo de arteriopatía coronaria), o *3)* para todas las demás investigaciones futuras que aprueben un CEIC y un comité de revisión científica. Para abordar los temas que surgen en el ejercicio b, podría también pedirse al participante el consentimiento específico para las investigaciones en las que podría secuenciarse su ADN. El participante puede estar de acuerdo con una, con dos o con todas las opciones. Por supuesto, es imposible describir las investigaciones futuras. De aquí que el consentimiento para futuras investigaciones no sea realmente informado, en el sentido de que el participante no conocerá la naturaleza, los riesgos y los beneficios de estudios futuros. Se pide al participante que confíe en que el CEIC y el comité de revisión científica solo permitirán estudios futuros que tengan solidez científica y ética.

2a. No dar al grupo testigo fármacos que se sabe que son eficaces sería perjudicial y, por lo tanto, no sería ético. Incluso si los participantes dieran su consentimiento informado a participar en un estudio controlado con placebo de ese tipo, un CEIC podría no aprobar el estudio porque viola los requisitos de las agencias de registro de que el equilibrio riesgo/beneficio sea aceptable y de que se reduzcan al mínimo los riesgos.

2b. Si todos los participantes en el estudio recibieran la quimioterapia que constituye el tratamiento estándar, también se podría aleatorizar a los participantes al nuevo tratamiento o al placebo. De manera alternativa, los investigadores podrían tratar de identificar un subgrupo de pacientes en los que ningún tratamiento ha demostrado prolongar la supervivencia (el criterio de valoración más significativo clínicamente en la mayor parte de los tratamientos del cáncer). Por ejemplo, a los pacientes cuya enfermedad ha progresado a pesar de varios tipos de quimioterapia estándar y que no tienen otras opciones que hayan sido eficaces, podría pedírseles que participaran en un ensayo controlado con placebo de la intervención experimental. Un grupo testigo aceptable podría ser el placebo o el mejor tratamiento actual. Este método supone que si el fármaco es activo en pacientes no tratados anteriormente, también lo será después de que otros tratamientos hayan fallado. Es, por supuesto, posible que un fármaco que no funcione en la enfermedad resistente al tratamiento pueda ser eficaz como tratamiento de primera línea.

3a. Durante el consentimiento informado, los investigadores deben comentar: *1)* la naturaleza del estudio; *2)* el número y la duración de las consultas; *3)* los posibles beneficios y riesgos de la participación (en este caso, fundamentalmente, los estigmas y la discriminación, si se viola la confidencialidad); *4)* alternativas a la participación en el estudio, entre ellas medidas de prevención de la infección por el VIH disponibles fuera del ensayo; *5)* la naturaleza voluntaria de la participación y el derecho a abandonar en cualquier momento, y *6)* la protección de la confidencialidad compatible con los requisitos de notificación exigidos por las autoridades de salud pública estatales.

3b. Los investigadores necesitan presentar la información de un modo que los participantes puedan entender. Los participantes con poca formación sanitaria no podrán entender un formulario de consentimiento informado por escrito. Sería útil que los investigadores consultaran con grupos de apoyo cómo presentar la información. Las sugerencias podrían ser videograbaciones, DVD y revistas de tiras cómicas. Deben realizarse pruebas preliminares. Además, los investigadores deben determinar qué malentendidos sobre el estudio suelen ser frecuentes, y revisar el proceso de consentimiento para evitarlos.

3c. Aun cuando el estudio sea un estudio de observación, los investigadores tienen una obligación ética de proporcionar información a los participantes sobre cómo reducir el riesgo de sufrir infección por el VIH. Hay razones éticas y científicas para hacerlo. Los investigadores tienen la obligación ética de evitar dañar a los participantes en su estudio. No pueden no aplicar medidas de salud pública disponibles que se sabe que evitan la enfermedad mortal, que es el criterio de valoración del estudio. Estas medidas incluirían asesoramiento, preservativos y envío a un programa de tratamiento de consumo de sustancias tóxicas y de cambio de agujas. Los investigadores pueden también invocar estas medidas para evitar dañar a los participantes en el siguiente ensayo de la vacuna, aun cuando la potencia del ensayo habrá disminuido.

Capítulo 15 Diseño de cuestionarios, entrevistas y encuestas por Internet

1a. No existe definición del tamaño de una «bebida».

1b. No hay modo de responder si el participante está bebiendo más de ocho bebidas al día.

1c. La pregunta no especifica el tiempo: días de la semana frente a fines de semana, todos los días frente a menos de diariamente.

1d. Puede que sea mejor especificar una estructura temporal concreta (p. ej., en los últimos 7 días).

2a. ¿Cuál de las siguientes afirmaciones describe mejor con qué frecuencia ha consumido bebidas alcohólicas durante el pasado año? Una bebida alcohólica incluye vino, licor o combinados.

Elija una de las ocho categorías:

○ Todos los días ○ 2-3 veces al mes

○ 5-6 días a la semana ○ Alrededor de una vez al mes

○ 3-4 días a la semana ○ Menos de 12 veces al año

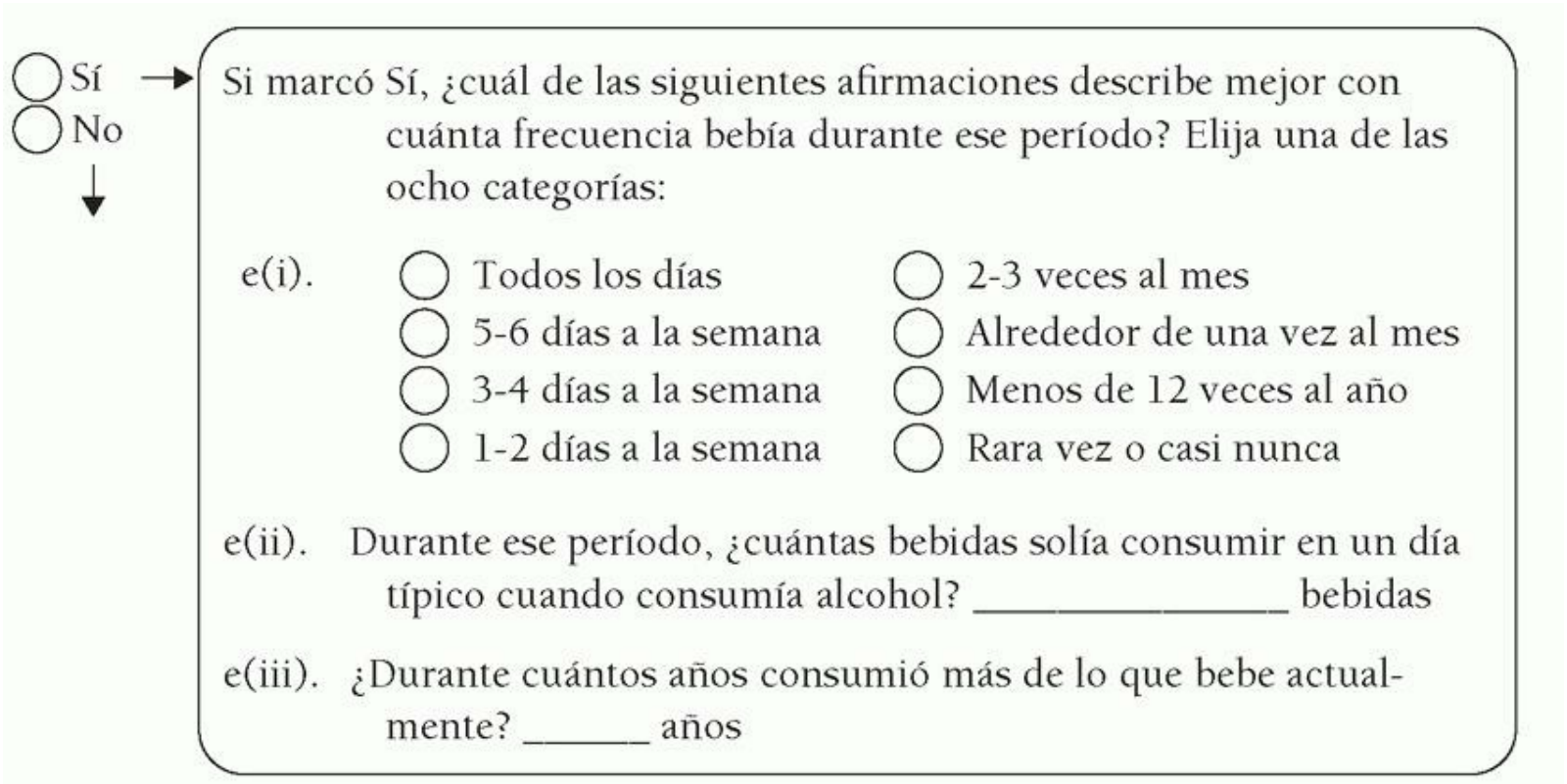
○ 1-2 días a la semana ○ Rara vez o casi nunca

2b. Durante el pasado año, ¿cuántas bebidas solía tomar en un día típico cuando consumía alcohol? Una bebida es alrededor de 350 cc de cerveza, 150 cc de vino o 45 cc de licor. _____ bebidas.

2c. Durante el pasado año, ¿cuál es el mayor número de bebidas alcohólicas que puede recordar haber consumido durante un día? _____ bebidas.

2d. ¿Qué edad tenía, aproximadamente, cuando empezó a consumir bebidas alcohólicas? _____ años de edad. (Si nunca ha consumido bebidas alcohólicas, escriba «nunca»)

2e. ¿Hubo un período en el que consumió más de lo que consume ahora?



Herramientas de imágenes

2f. ¿Ha tenido alguna vez lo que se podría considerar un problema con la bebida?

○ Sí

○ No

3a. La obtención de datos mediante entrevistas necesita más formación y tiempo del personal que un cuestionario autoadministrado, y es, por lo tanto, más cara.

3b. A algunos participantes no les gusta responder a otra persona sobre cuestiones delicadas acerca del comportamiento sexual.

3c. Salvo que los entrevistadores estén bien preparados y las entrevistas estén normalizadas, la información obtenida puede variar.

3d. Sin embargo, los entrevistadores pueden repetir y sondear de un modo que mejore la comprensión y produzca unas respuestas más exactas y completas, en algunas situaciones, que un cuestionario autoadministrado.

Capítulo 16 Gestión de los datos

1.

ID Particip.	AlgunaVez Fumó100cig	EdadPrimer Cigar	PromCigDía	AlgúnCig SemanaPasada	CigarDia SemanaPasada	EdadDejó Fumar
1 001	1	17	30	1	30	
1 002	1	21	20	0		45
1 003	0			0		

Así podrían aparecer los datos en un programa de hoja de cálculo como Excel. Hay muchas posibilidades aceptables para los nombres de campo (encabezamientos de las columnas). Estos nombres de campo usan letras mayúsculas en mitad del término para separar sus partes. Los diseñadores de bases de datos se dividen por igual entre aquellos a los que les gusta este sistema de uso de mayúsculas y aquellos a los que no.

2a.

IDPartic	NúmKit	FechaAdm	HorAdm	FinCPre Hosp	HFinCPre Hosp	LlegHosp	ActCLleg Hosp	ECGVLleg Hosp
189	A322	3/12/1994	17:39	0		17:48	1	
410	B536	12/1/1998	01:35	1	01:39	01:53	0	4

2b.

Nombre del campo	Tipo de dato	Descripción	Norma de validación
IDPartic	Entero	Identificador de participante único	
NúmKit	Texto(5)	Código farmacéutico de investigación de cinco caracteres	

FechaAdm	Fecha	Fecha de administración del fármaco del estudio	
HorAdm	Hora	Hora de administración del fármaco del estudio	
FinCPreHosp	Sí/no	¿Cesó la convulsión antes de llegar al hospital?	
HFinCPreHosp	Hora	Hora en que cesaron las convulsiones antes de llegar al hospital (en blanco si la convulsión no cesó)	
LlegHosp	Hora	Hora de llegada al hospital	
ActCLlegHosp	Sí/no	¿Siguió la actividad convulsiva al llegar al hospital?	Comprobar con FinCPreHosp
ECGVLlegHosp	Entero	ECG verbal al llegar al hospital (dejar en blanco si seguían las convulsiones)	Entre 1 y 5

2c. Ventajas de un formulario en pantalla:

- No se necesita transcripción desde formularios en papel a las tablas de datos informatizadas
- Retroalimentación inmediata sobre entradas no válidas
- Lógica de exclusión programada (si las convulsiones cesaron antes de llegar al hospital, el formulario informático señala la hora en que cesaron; si no, este campo se inhabilita y se salta)
- Puede estar disponible a través de un explorador de web en muchos sitios al mismo tiempo

Inconvenientes de un formulario en pantalla:

- Se necesita equipo: una estación de trabajo informática
- Se necesita formación del usuario

Ventajas de un formulario en papel:

- Fácil y rápido de usar
- Puede transportarse
- Posibilidad de introducir información imprevista o datos no estructurados (notas al margen, respuestas que, de otro modo, no se considerarían, etc.)
- Material necesario: un bolígrafo
- La formación la recibe todo el personal de introducción de datos en la formación básica

Inconvenientes de un formulario en papel:

- Necesita la posterior transcripción a una base de datos informática
- No hay retroalimentación interactiva ni lógica de exclusión
- La observación y la introducción de datos se limitan a una persona por centro

Aunque la introducción de datos con formularios de recopilación de datos en pantalla tiene muchas ventajas y la recomendamos en la mayor parte de los estudios de investigación, en este estudio no es práctica. La forma más sencilla, rápida y agradable para el usuario de capturar datos en un medio no volátil sigue siendo el uso de bolígrafo y papel.

3. Cuando se codifica con 0 para *no* o *ausente*, y con 1 para *sí* o *presente*, el valor promedio de una variable dicotómica (sí/no) se puede interpretar como la proporción con el atributo. De los pacientes a los que se aleatorizó a recibir lorazepam, el 40,9 % (27 de 66) seguía con convulsiones al llegar al hospital; de los pacientes a los que se aleatorizó a recibir diazepam, el 57,4 % (39 de 68) seguía con convulsiones; y de los pacientes a los que se aleatorizó a recibir placebo, el 78,9 % (56 de 71) seguía con convulsiones.

Capítulo 17 Realización del estudio y control de calidad

1a. ¡No es suficiente! Pero hay algunos pasos que puede dar:

- Identificar todos los valores ausentes y fuera del intervalo, y volver a comprobar los formularios en papel para asegurarse de que se introdujeron los datos correctamente.
- Recuperar datos perdidos con las historias.
- Recopilar datos de entrevistas perdidos de pacientes supervivientes (aunque esto no ayudará con los que fallecieron ni con aquellos cuyas respuestas pudieran haber cambiado con el tiempo).
- Hacer un esfuerzo especial para encontrar participantes que se hayan perdido durante el seguimiento, y conseguir, al menos, una entrevista telefónica con ellos.
- Conseguir la situación vital, a través del *National Death Index* o alguna compañía que busque personas.

1b. Recopilar menos datos:

- Comprobar los formularios sobre el terreno inmediatamente después de recopilar los datos para asegurarse de que se han cumplimentado todas las preguntas, y de que son exactas.
- Usar una introducción de datos interactiva con comprobaciones de valores ausentes, fuera de los límites válidos e ilógicos.
- Revisar brevemente la base de datos tras la introducción de los datos, de forma que puedan conseguirse los datos ausentes antes de que el paciente se vaya del hospital (o fallezca).
- Tabular periódicamente las distribuciones de los valores de todas las preguntas durante el estudio, para identificar valores ausentes, valores fuera del intervalo y posibles errores.
- Mantener reuniones periódicas con el equipo para revisar el progreso e insistir en la importancia de disponer de datos completos.

Capítulo 18 Estudios comunitarios e internacionales

1a. La consulta de aparato digestivo:

- Ventajas: probablemente sea una fuente conveniente y accesible de pacientes. El personal de la consulta probablemente tenga experiencia en participar en una investigación. No debería resultar difícil realizar una serie habitual de pruebas diagnósticas a los pacientes con dolor abdominal.
- Inconvenientes: los pacientes pueden ser un subgrupo muy seleccionado de pacientes de la comunidad que acuden con dolor abdominal, y la evolución clínica de estos pacientes puede diferir de la de otros de la comunidad. Por lo tanto, los resultados pueden tener una posibilidad de generalización escasa.

1b. Consultas de la comunidad:

- Ventajas: aquí puede identificar pacientes en la primera presentación sin la selección ni la demora causados por el proceso de derivación. Los médicos de atención primaria pueden beneficiarse de la oportunidad de participar en la investigación.
- Inconvenientes: son fundamentalmente logísticas. Identificar médicos y pacientes que participen, y realizar un protocolo de investigación estándar, serán importantes tareas de organización, y el control de calidad puede ser un reto.

2a. Solo puede responderse con datos locales. La investigación de otros lugares no ayudará.

2b. Es algo muy conocido por la bibliografía internacional. No es probable que repetir una investigación de este tipo en China sea un uso eficaz de los recursos.

2c. Para esta pregunta, es probable que la posibilidad de generalización de la investigación a otros lugares sea intermedia. Las estrategias para dejar de fumar que fueron útiles en otros países pueden servir como base para las estrategias a estudiar en China, pero no se puede estar seguro de que tengan el mismo éxito en China sin una investigación local. Pueden ser útiles estudios anteriores en poblaciones de otros lugares con lazos culturales con China, como inmigrantes chinos recientes a Estados Unidos.

Capítulo 19 Elaboración de una propuesta para la financiación de la investigación

1-3. Esperamos que haya tenido algunas ideas útiles para planificar su propia agenda de investigación, y le animamos a que implique a sus mentores y sus compañeros en la discusión de la mejor manera de continuar.

Glosario
NA

Abandono. Participante en un estudio en el que no se puede evaluar el estado de la variable de respuesta, muchas veces porque se ha negado a continuar el seguimiento. En ocasiones esto incluye a los participantes que abandonaron porque murieron durante el estudio. Por ejemplo, en un estudio hubo 17 abandonos: ocho por negativa a seguir, seis por muerte y tres porque presentaron demencia.

Ajuste. Nombre general para diversas técnicas estadísticas utilizadas para tener en consideración los efectos de una o más variables, o una asociación entre otras dos variables. Por ejemplo, el ajuste del nivel de ingresos redujo la magnitud de la asociación entre el nivel educativo y la mortalidad.

Ajuste excesivo. Problema que surge cuando los investigadores seleccionan variables o valores de corte para un modelo multivariado basados en parte en la variación aleatoria de la muestra, lo que lleva a una escasa posibilidad de generalizar los resultados. Por ejemplo, los revisores sospecharon un ajuste excesivo cuando los autores describieron que el mejor modelo para predecir las cataratas recurrentes incluía el nacimiento en los meses de marzo o agosto en mujeres de entre 65 y 74 años de edad.

Ajuste multivariado. Término general para las técnicas estadísticas utilizadas para ajustar los efectos de una o más posibles variables de confusión sobre la asociación entre una variable predictiva y otra de respuesta. Por ejemplo, utilizando ajuste multivariado, en un estudio se encontró que la ingesta de suplementos de vitamina D se asociaba a aumento del riesgo de deterioro cognitivo después de ajustar la edad, el sexo, el nivel educativo, la función cognitiva inicial y el tabaquismo.

Aleatorización. Proceso de asignar aleatoriamente a los participantes elegibles a uno de los grupos del estudio en un ensayo aleatorizado. El número de grupos de tratamiento y la probabilidad de ser asignado a cada uno de los grupos se determinan antes del comienzo de la aleatorización. Aunque habitualmente se asigna a los participantes elegibles a dos grupos del estudio con la misma probabilidad (50 %), se puede realizar la asignación aleatoria a cualquier número de grupos en estudio, con cualquier probabilidad predeterminada. Por ejemplo, en un estudio en el que se comparen dos tratamientos con un testigo con placebo, se podría realizar la aleatorización a tres grupos, con asignación del 30 % a cada uno de los dos grupos de tratamiento activo y del 40 % al grupo del placebo.

Aleatorización por bloques. Método para asignar a los participantes a una intervención particular por bloques (grupos) de un tamaño preespecificado (p. ej., cuatro o seis) para garantizar que se asigne a números similares de participantes a los grupos de intervención y testigo. A menudo se utiliza en estudios multicéntricos en los que los investigadores quieren que los números totales de participantes que reciben la intervención y los testigos sean similares en cada uno de los centros. Por ejemplo, se asignó aleatoriamente a los pacientes de cada consulta a los grupos de tratamiento o testigo en bloques de seis, asegurándose de que el número de participantes por grupo difiriera en no más de tres. Véase también *Aleatorización por bloques estratificada*.

Aleatorización por bloques estratificada. Procedimiento de aleatorización diseñado para garantizar que se asigne aleatoriamente a cada uno de los grupos del estudio a números iguales de participantes con una característica determinada (habitualmente un factor de confusión). La aleatorización se estratifica por la característica de interés; en cada uno de los estratos se asigna aleatoriamente a los participantes a bloques de un tamaño predefinido. Por ejemplo, en un ensayo de un fármaco para prevenir las fracturas, el antecedente de fractura vertebral es un factor predictivo tan potente del resultado y de la respuesta a muchos tratamientos que sería mejor asegurarse de que hubiera un número igual de participantes con y sin fractura vertebral en cada uno de los grupos del estudio. Por lo tanto, los investigadores utilizaron aleatorización por bloques estratificada para dividir a los participantes en dos estratos (los que habían tenido fracturas vertebrales y los que no); dentro de cada uno de los estratos se realizó aleatorización por bloques de seis a 10 participantes.

Aleatorización por conglomerados. Técnica en la que se asigna aleatoriamente a grupos de participantes, conocidos como conglomerados, a diferentes tratamientos, en lugar de asignar aleatoriamente a cada uno de los participantes de manera individual. Por ejemplo, en un estudio de los efectos de la reducción del ruido sobre la recuperación de la cirugía cardíaca, el investigador utilizó aleatorización por conglomerados para asignar a las unidades de cuidados intensivos de 40 hospitales diferentes a una intervención «postoperatoria tranquila» o a una situación testigo de «asistencia habitual».

Aleatorización mendeliana. Técnica para mejorar la inferencia causal aprovechando la herencia aleatoria de genes que afectan a la susceptibilidad a un factor de riesgo o un tratamiento. Por ejemplo, la probabilidad de una relación causal entre el uso de paracetamol por la madre y el asma de los hijos mejoró por la observación de que la asociación era significativamente mayor en madres con el genotipo T1 de la glutatión S-transferasa, una enzima implicada en la destoxificación de un metabolito del paracetamol.

Alfa (α). Cuando se diseña un estudio, la probabilidad máxima preestablecida de cometer un error de tipo I, es decir, de rechazar la hipótesis nula cuando es verdadera. Por ejemplo, al elegir un valor de a de 0,05, el investigador estableció una probabilidad máxima del 5 % de que su estudio encontrara una asociación estadísticamente significativa entre la raza no blanca y el riesgo de cáncer de colon sólo por el azar. También se denomina nivel de significación estadística.

Análisis de los datos secundarios. Utilización de datos existentes para investigar preguntas de la investigación diferentes a aquellas para las que se recogieron originalmente los datos. Los conjuntos de datos secundarios pueden incluir estudios de investigación previos, historias clínicas, datos de facturación de asistencia sanitaria y certificados de defunción. Por ejemplo, se podían utilizar los datos de las altas hospitalarias y del registro de mortalidad nacional para el análisis de los datos secundarios a fin de determinar la mortalidad al cabo de 1 año en pacientes con diagnóstico al alta de pancreatitis aguda.

Análisis de intención de tratar. En un ensayo aleatorizado, proceso de comparar los participantes de acuerdo con el grupo al que se les asignó aleatoriamente, aunque no sea el mismo que el tratamiento que han recibido realmente. Se trata de la forma más rigurosa de análisis. Por ejemplo, los investigadores realizaron un análisis de intención de tratar para determinar si la asignación aleatoria a recibir 6 meses de psicoterapia mejoraba los síntomas de ansiedad en comparación con la asignación aleatoria a un grupo testigo que recibió un folleto sobre reducción del estrés. Véase también *Análisis según el protocolo*.

Análisis según el protocolo. En un ensayo clínico, abordaje en el cual se incluyen los datos de los participantes sólo si los participantes cumplieron el protocolo del estudio, lo cual se define generalmente como haber tomado o utilizado la intervención del estudio tal y como se les indicó. Por ejemplo, en un ensayo aleatorizado el que se comparó la cirugía con la fisioterapia para el tratamiento de la artrosis grave de la rodilla, en un análisis según el protocolo se incluyeron los datos únicamente de los participantes del grupo de cirugía a los que realmente se operó, y los de los participantes del grupo de fisioterapia que cumplieron el régimen de fisioterapia. Véase también *Análisis de intención de tratar*.

Análisis de la sensibilidad. Utilización de diferentes métodos (p. ej., definiciones alternativas de las variables predictivas o de respuesta, diferentes pruebas estadísticas) para determinar si los resultados del análisis principal son sólidos. Por ejemplo, en un metaanálisis de ensayos clínicos del efecto de los inhibidores selectivos de la recaptación de serotonina sobre la depresión, en un análisis de la sensibilidad el investigador podría incluir sólo los estudios con enmascaramiento para demostrar que los resultados son sólidos cuando el análisis se restringe a ensayos de calidad elevada.

Análisis de subgrupos. Comparaciones entre grupos asignados aleatoriamente en un subconjunto de los participantes de un ensayo. Por ejemplo, en un ensayo aleatorizado del efecto de un modulador selectivo del receptor estrogénico (MSRE) sobre la recurrencia del cáncer de mama, los investigadores realizaron un análisis de subgrupos del efecto del tratamiento por estadio del cáncer, comparando el efecto del SERM con el placebo en mujeres con enfermedad en estadio I, II, y III y IV.

Análisis de supervivencia. Técnica estadística que se utiliza para comparar el tiempo transcurrido hasta una variable de respuesta (no necesariamente la supervivencia) entre los grupos de un estudio. Por ejemplo, en un ensayo aleatorizado del efecto de la cirugía de derivación arterial coronaria en comparación con la angioplastia coronaria percutánea para la prevención del infarto de miocardio y la muerte, se podría utilizar el análisis de supervivencia para comparar el tiempo desde el tratamiento inicial hasta cualquiera de esas variables de respuesta de los dos grupos.

Árboles de clasificación y regresión (ACR). Véase *Partición recursiva*.

Asociación. Relación cuantificable entre dos variables. Por ejemplo, en el estudio se encontró una asociación entre el sexo masculino y el riesgo de deterioro cognitivo en personas de 60 a 69 años de edad, con un cociente de riesgos instantáneos de 1,6.

Asociación espuria. Asociación entre una variable predictiva y una variable de respuesta que se observa en un estudio, pero que no es cierta en la población, ya sea por azar o sesgo. Por ejemplo, en estudios observacionales se encontró disminución del riesgo de

enfermedad cardiovascular en personas que tomaban suplementos de β-caroteno. Sin embargo, en un ensayo aleatorizado de suplementos de β-caroteno no se encontró ningún efecto sobre el riesgo de enfermedad cardiovascular, lo que indica que la asociación observada en los estudios de observación era espuria.

Base de datos relacional. Programa informático que permite el almacenamiento de información relacionada en una serie de tablas. Las tablas pueden estar vinculadas entre sí por campos comunes. Por ejemplo, una base de datos relacional para un estudio podría incluir la IDEstudio y la FechaNacimiento de cada uno de los participantes en una tabla de Participantes, y la IDEstudio y la FechaVisita en una tabla de Encuentros, que podría tener muchos encuentros por participante. La edad de un participante el día de un encuentro se puede calcular fácilmente utilizando la IDEstudio para relacionar cada una de las FechaVisita con la fecha de nacimiento del participante.

Beta (β). Cuando se diseña un estudio, la probabilidad máxima preestablecida de cometer un error de tipo II, es decir, de no rechazar la hipótesis nula cuando es falsa. Esta medida sólo es significativa en el contexto de una magnitud del efecto determinada. Por ejemplo, si un investigador especifica un valor de β de 0,20 (y de α de 0,05), necesitaría aproximadamente 25 000 participantes por grupo con un seguimiento de 10 años para demostrar que el uso diario de ácido acetilsalicílico reduce a la mitad el riesgo de cáncer de colon. Planteado de otra manera, si el ácido acetilsalicílico tuviera realmente ese efecto, un estudio de 25 000 participantes por grupo tendría una probabilidad del 20 % de no rechazar la hipótesis nula de ausencia de diferencia (con a = 0,05). Véase también *Potencia*.

Búsqueda. Comando o instrucción en una base de datos relacional para seleccionar o manipular los datos. Por ejemplo, el coordinador del estudio escribió una búsqueda para seleccionar los nombres y la información de contacto de todos los participantes en el estudio que debían acudir a una visita de seguimiento en los 2 meses siguientes y para los que todavía no se había programado la cita.

Calibración. Proceso de garantizar que un instrumento ofrezca una lectura constante; habitualmente se realiza midiendo un patrón conocido y después ajustando (calibrando) el instrumento en consecuencia. Por ejemplo, la balanza se calibró todos los meses pensando un bloque de acero de 50 kg.

Campo. Columna de una tabla de una base de datos relacional que incluye datos sobre un atributo específico del registro. Por ejemplo, dos de los campos de la tabla Encuentros serán IdPartic (para relacionarlo con la información específica del participante) y PesoKg (peso en kg).

Caso. Participante que tiene, o que llega a tener, la variable de respuesta de interés. Por ejemplo, se definieron los casos como los que tuvieron angina inestable, infarto de miocardio o muerte súbita durante el seguimiento. Véase también *Testigo*.

Causa-efecto. Concepto de que una variable predictiva es responsable de la producción de una variable de respuesta, o aumenta la probabilidad de aparición de la variable de respuesta. La finalidad de la mayoría de los estudios de observación es demostrar una relación de causa-efecto, aunque esto es difícil, salvo que la causa (p. ej., un tratamiento) se asigne aleatoriamente. Por ejemplo, el investigador realizó un estudio de casos y testigos para determinar si había una relación de causa-efecto entre el consumo de alcohol (la causa) y el cáncer pancreático (el efecto). Véanse también *Confusión* y *Efecto-causa*.

Clave principal. En una base de datos relacional, el campo o la combinación de campos que identifican de manera única a cada fila de una tabla particular. Por ejemplo, el investigador creó un NúmeroDeVisita único para que fuera la clave principal de una tabla de visitas ambulatorias.

Cociente de riesgos. El cociente de la tasa de riesgos instantáneos en las personas expuestas a un factor de riesgo dividido por la tasa de riesgos instantáneos en las personas no expuestas; casi siempre se estima con un modelo de riesgos proporcionales (modelo de Cox). Por ejemplo, el cociente de riesgos de presentar arteriopatía coronaria era de 2,0 cuando se comparó a hombres de 50 a 59 años de edad con mujeres de la misma edad.

Coeficiente de correlación. Término estadístico que indica el grado en el que las mediciones continuas están relacionadas linealmente, de manera que el cambio de una medición se asocia a un cambio proporcional de la otra. A menudo se abrevia como *r*. Por ejemplo, en una muestra de mujeres de mediana edad, la altura y el peso se correlacionaban con un valor de $r = 0,7$.

Coeficiente de variación (CV). Medida de la precisión de una medición, que se obtiene dividiendo la desviación típica de una serie de mediciones realizadas en una única muestra por la media de dichas mediciones. En ocasiones se calcula el CV de los valores de la porción central y los extremos de la medición. Por ejemplo, el laboratorio determinó que el coeficiente de variación de la concentración sérica de estradiol era del 10 % en una muestra de mujeres perimenopáusicas (en las que en la concentración de estradiol era muy baja), y de tan sólo el 2 % en mujeres más jóvenes.

Cointervención. En un ensayo clínico, intervención que se realiza después de la aleatorización, distinta a la intervención que se estudia, que afecta a la probabilidad de una variable de respuesta. Las cointervenciones que se producen con incidencias distintas en los grupos del estudio pueden sesgar la variable de respuesta y hacer que sea difícil atribuir la causalidad a la intervención que se estudia. Por

ejemplo, fue difícil interpretar un estudio del efecto de una intervención para fomentar la lactancia materna sobre la enfermedad alérgica posterior en lactantes porque las mujeres del grupo de intervención no sólo dieron de mamar más tiempo, sino que también tuvieron más probabilidad que el grupo testigo de retrasar la introducción de alimentos sólidos y de comprar fórmulas hipoalergénicas, situaciones ambas que son posibles cointervenciones.

Comité directivo. En un estudio multicéntrico, comité que regula los aspectos generales del estudio. Generalmente está formado por los investigadores principales de todos los centros de investigación, el centro coordinador y representantes del promotor. Por ejemplo, el comité directivo del estudio decidió si se debían realizar o no los estudios auxiliares propuestos.

Concordancia. Medida del acuerdo entre dos (o más) observadores sobre la aparición de un fenómeno. Por ejemplo, la concordancia entre los radiólogos A y B era del 96 % para la presencia de un infiltrado pulmonar lobular, aunque era de tan sólo el 76 % para la cardiomegalia. Véase también *Kappa*.

Condicionamiento. Proceso de examinar las asociaciones entre dos o más variables distintas con niveles fijos de la variable «a la que se condiciona». La especificación, el emparejamiento, la estratificación y el ajuste multivariado son las formas más frecuentes de realizar el condicionamiento a una variable. Por ejemplo, el investigador no encontró ninguna asociación entre el consumo de cocaína y el riesgo de sífilis después de condicionarlo al número de parejas sexuales.

Condicionamiento a un afecto compartido. Fuente de sesgo en estudios epidemiológicos en los que se introduce una asociación entre dos causas diferentes del mismo efecto mediante el condicionamiento a dicho efecto. Por ejemplo, debido al condicionamiento a un afecto compartido (tiempo de pantalla total), hay una asociación inversa entre la televisión y el juego en niños que pasaban al menos 6 h al día delante de la pantalla: los que pasaban más tiempo viendo la televisión, pasaban menos tiempo jugando a videojuegos.

Conducta científica impropia. Término general para la situación en la que se defrauda intencionadamente a la comunidad científica, que incluye la conducta de investigación impropia (invención y falsificación de los datos o plagio), además de la autoría invitada y «fantasma», y el conflicto de interés que no se comunica ni se resuelve. Por ejemplo, la institución del investigador consideró que era culpable de conducta científica impropia porque no había comunicado su interés patrimonial en la compañía que fabricaba el dispositivo médico que estaba estudiando.

Conducta de la investigación impropia. Conducta ilegal o no ética en relación con la investigación, como plagio e invención o falsificación de los datos de la investigación. Por ejemplo, se encontró que un coordinador de investigación del VA Medical Center de Albany, Nueva York, había enviado de manera repetida documentación falsa para permitir la inclusión de personas que no cumplían los criterios de participación. Posteriormente se excluyeron todos los datos de Albany, de modo que se desperdició el tiempo y el esfuerzo de los participantes. Véase también *Conducta científica impropia*.

Confusión. Fenómeno epidemiológico en el que una asociación entre una variable predictiva y una variable de respuesta se debe a una tercera variable (denominada factor de confusión o variable de confusión), en lugar de ser una relación de causa-efecto entre la variable predictiva y la variable de respuesta. Por ejemplo, en la aparente asociación entre el tabaquismo y el cáncer cervical se introdujo confusión por la infección por el virus del papiloma humano (VPH), porque las mujeres que fuman también tienen más probabilidades de tener (múltiples parejas sexuales e) infección por el VPH. Véase también *Modificación del efecto*.

Confusión por indicación. Forma específica de confusión en la que una de las indicaciones de un tratamiento es el factor de confusión; habitualmente se produce en estudios de observación de la asociación entre un tratamiento y una variable de respuesta. Por ejemplo, los revisores de un estudio de observación estaban preocupados de que la asociación descrita entre un nuevo tratamiento del trastorno bipolar y el aumento del riesgo de suicidio pudiera deberse a que los pacientes con una enfermedad subyacente más grave habían sido tratados selectivamente con el nuevo medicamento.

Contaminación. Proceso indeseable por el cual alguno o la mayoría de los efectos de una intervención también afectan a los participantes del grupo testigo. Por ejemplo, un estudio de los efectos de si enseñar a los niños a contar hacia atrás mejoraba sus habilidades aritméticas totales sufrió contaminación porque los niños del grupo de la intervención no pudieron resistirse a enseñar esa habilidad a sus amigos del grupo testigo.

Control de calidad. Proceso para garantizar que la realización de un estudio, con pasos como la inclusión, las mediciones, los procedimientos de laboratorio, y la gestión y el análisis de los datos, tenga la máxima calidad. Por ejemplo, los investigadores controlaron la calidad de la recogida de los datos preparando procedimientos por escrito explícitos para todas las mediciones del estudio en un manual de instrucciones, y observando intermitentemente al personal del estudio para asegurarse de que seguía dichos procedimientos.

Corrección de Bonferroni. Técnica para evitar el error de tipo I dividiendo el valor de a total de un estudio por el número de hipótesis estudiadas. Por ejemplo, como los investigadores estaban estudiando cuatro hipótesis diferentes, utilizaron la corrección de Bonferroni

para reducir el valor de a de cada una de las hipótesis desde 0,05 hasta 0,0125.

Criterios de exclusión. Lista de atributos que impiden que un posible participante sea elegible para un estudio. Por ejemplo, los criterios de exclusión del estudio fueron tratamiento previo con un antidepresivo en los 2 años previos, uso actual de a-bloqueantes o β-bloqueantes, e imposibilidad de leer inglés a nivel de sexto curso. Véase también *Criterios de inclusión*.

Criterios de inclusión. Lista de atributos que deben cumplir los posibles participantes en un estudio. Por ejemplo, los criterios de inclusión en un estudio fueron personas de 18 a 65 años de edad que vivían en San Francisco y no tenían antecedentes de depresión. Véase también *Criterios de exclusión*.

Criterios de participación. Lista de atributos que deben tener los pacientes para poder participar en un estudio. Los criterios de selección pueden variar si se incluye a los participantes en diferentes grupos, como en estudios de casos y testigos o de dobles cohortes. Por ejemplo, los criterios de selección para un estudio de un nuevo tratamiento de la gota incluyeron edad entre 20 y 75 años, al menos un episodio de gota diagnosticado por un médico en los 12 meses previos, y una concentración sérica de ácido úrico de al menos 6 mg/dl. Véanse también *Criterios de exclusión* y *Criterios de inclusión*.

Criterios de selección. Reglas que definen quién es idóneo para participar en un estudio, y que abarcan los criterios de inclusión y exclusión. Por ejemplo, en un ensayo clínico de testosterona transdérmica para aumentar la libido en mujeres posmenopáusicas, los criterios de selección podrían ser mujeres de 45 a 60 años de edad con libido baja que no tengan cardiopatía isquémica y que no hayan tenido más de tres períodos menstruales en el año previo.

Cruzamiento. Término utilizado para describir a un participante, habitualmente en un ensayo clínico, que comienza en un grupo (p. ej., la asistencia habitual) y cambia al otro grupo (p. ej., el tratamiento activo) durante el estudio. La mayoría de las veces se produce cuando el tratamiento activo supone un procedimiento. Por ejemplo, durante el estudio se cruzó a 15 participantes con cáncer de próstata a los que se asignó inicialmente a espera vigilante a recibir radioterapia o cirugía.

Cuestionario. Instrumento de medición formado por una serie de preguntas para obtener información de los participantes del estudio. Los cuestionarios puede ser autoadministrados o administrados por el personal del estudio. Por ejemplo, en el cuestionario de frecuencia de alimentos por bloques *(Block Food Frequency Questionnaire)* se pregunta por la ingesta habitual de 110 alimentos distintos para evaluar la ingesta de múltiples nutrientes y grupos de alimentos.

Curva de eficacia diagnóstica (ROC). Técnica gráfica para cuantificar la exactitud de una prueba diagnóstica e ilustrar el equilibrio entre la sensibilidad y la especificidad con diferentes valores liminares que establecen la positividad de la prueba. La curva muestra la incidencia de valores positivos verdaderos (sensibilidad) en el eje de ordenadas y la correspondiente tasa de falsos positivos (1 - especificidad) en el eje de abscisas para diversos valores de corte que definen la positividad de la prueba. El área bajo la curva ROC, que varía de 0,5, para una prueba inútil, hasta 1,0 para una prueba perfecta, es un resumen útil de la exactitud general de la prueba. Por ejemplo, el área bajo la curva ROC para el uso de la TC (que se podría interpretar como claramente positiva, probablemente positiva, no útil, probablemente normal o claramente normal) para diagnosticar apendicitis fue 0,95, mucho mejor que el valor de 0,77 para la ecografía (que tuvo categorías de interpretación similares).

Datos. Nombre en plural que se utiliza para describir mediciones, habitualmente en formato numérico. Por ejemplo, los datos sobre la prevalencia de diversas enfermedades son útiles cuando se toman decisiones sobre la asignación de los recursos sanitarios.

Datos ausentes. Datos que no se obtuvieron durante un estudio, ya sea en situación inicial o durante el seguimiento. Por ejemplo, el investigador estaba preocupado de que la proporción relativamente grande (34 %) de pacientes de los que no se disponía de datos de consumo de alcohol pudiera haber sesgado su estudio sobre los factores de riesgo de caídas.

Desviación típica. Medida de la varianza (dispersión) de una variable continua. Por ejemplo, el investigador señaló que la media de edad de una cohorte de 450 hombres era de 59 años, con una desviación típica de 10 años.

Diccionario de los datos. Tabla u hoja de cálculo que incluye información sobre cada una de las variables de un estudio, como su nombre y tipo (p. ej., numérico o alfanumérico), la definición de cada una de las variables y el intervalo de valores permitido. Por ejemplo, el investigador consultó el diccionario de los datos porque había olvidado que un «5» en el campo llamado «raza» se utilizaba para indicar indio estadounidense/nativo de Alaska.

Diferencia de riesgos. Riesgo de una variable de respuesta en un grupo menos el riesgo en un grupo de comparación. Por ejemplo, si el riesgo de episodios tromboembólicos venosos en mujeres que utilizan habitualmente estrógenos es 5/1 000 (0,5 %) y el riesgo en las que nunca han utilizado estrógenos es 2/1 000 (0,2 %), la diferencia de riesgos en mujeres que utilizan estrógenos en comparación con las que no los utilizan es 3/1 000 (0,3 %). Véase también *Número que es necesario tratar*.

Diseño entre grupos. Diseño de un estudio que compara las características o variables de respuesta de los participantes en dos (o más) grupos diferentes. Por ejemplo, el investigador utilizó un diseño entre grupos para comparar las tasas de mortalidad intrahospitalaria en pacientes tratados en unidades de cuidados intensivos que tenían intensivistas de manera continua con las de aquellos pacientes tratados en unidades que utilizaban monitorización electrónica de los pacientes. Véase también *Diseño intragrupo*.

Diseño intragrupo. Diseño de estudio en el que se comparan las mediciones en un único grupo de participantes, la mayoría de las veces en dos períodos temporales diferentes. Este diseño elimina la confusión, porque cada participante es su propio testigo. Sin embargo, los diseños intragrupo son susceptibles de efectos de aprendizaje, regresión a la media y tendencias a largo plazo. Por ejemplo, utilizando un diseño intragrupo se midió la glucemia basal en un grupo de pacientes con diabetes antes de comenzar un programa de ejercicio y después de la finalización del programa para determinar si el ejercicio reducía la glucemia basal. Véanse también *Diseño entre grupos, Prueba de la* t *para una muestra* y *Diseño de series cronológicas.*

Diseño de series cronológicas. Diseño de estudio intragrupo en el que se realizan mediciones antes y después de que cada uno de los participantes (o una comunidad completa) reciba una intervención. Este diseño elimina la confusión, porque cada uno de los participantes es su propio testigo. Sin embargo, los diseños intragrupo son susceptibles a los efectos del aprendizaje, la regresión a la media y las tendencias a largo plazo. Por ejemplo, utilizando un diseño de series cronológicas se midió la glucemia basal en un grupo de pacientes con diabetes antes de comenzar un programa de ejercicio y de nuevo después de finalizar el programa para determinar si el ejercicio reducía la glucemia basal. Véase también *Diseño intragrupo*.

Dosis-respuesta. Fenómeno por el cual cuanto mayor sea la exposición (dosis), mayor será la magnitud o la probabilidad de la variable de respuesta (respuesta). (Si una exposición es protectora, entonces cuanto mayor sea la exposición, menor será la probabilidad de la variable de respuesta.) Por ejemplo, en un estudio se encontró una relación dosis-respuesta entre la exposición al sol y el número de nevos melanocíticos; en otros se encontró una relación dosis-respuesta entre el número de nevos y el riesgo de melanoma.

Efectividad. Aunque no hay ninguna definición estándar de este término, lo utilizamos para referirnos a una medida del grado en que una intervención actúa adecuadamente en la práctica real, en contraposición con la forma en la que actúa en un ensayo aleatorizado. Por ejemplo, como en varios ensayos clínicos realizados en contextos urbanos se ha encontrado que el activador del plasminógeno hístico (tPA) redujo la morbimortalidad cardiovascular, los investigadores estudiaron su efectividad en 25 servicios de urgencias rurales. Véase también *Eficacia*.

Efecto-causa. Situación en la que una variable de respuesta da lugar a la variable predictiva, en lugar de que ocurra lo contrario. Por ejemplo, aunque en un estudio de casos y testigos se observó que la exposición a los broncodilatadores inhalados se asociaba a un aumento del riesgo de neumopatía intersticial, la explicación más probable era de efecto-causa, es decir, que los pacientes con neumopatía intersticial tenían más probabilidad de haber sido tratados (erróneamente) con inhaladores. Véase también *Causa-efecto*.

Eficacia. Aunque no hay ninguna definición estándar de este término, lo utilizamos para referirnos a una medida del grado en que una intervención actúa adecuadamente en un ensayo clínico, en contraposición con cómo actuaría en la práctica real. Por ejemplo, en un ensayo clínico se describió que el activador del plasminógeno hístico (tPA) tuvo una eficacia del 25 % en la reducción de la morbimortalidad en pacientes con accidente cerebrovascular agudo. Véase también *Efectividad*.

Emparejamiento. En un estudio de casos y testigos, proceso de seleccionar los testigos para que sean similares en determinados atributos a los casos, para reducir la confusión por esos atributos. Por ejemplo, en un estudio de casos y testigos de los factores de riesgo de brucelosis se emparejó a los testigos con los casos por su edad (con una diferencia de menos de 3 años), el sexo y el país de residencia. Véase también *Emparejamiento excesivo*.

Emparejamiento excesivo. Situación en la que el emparejamiento más allá de lo necesario para controlar la confusión reduce la capacidad del investigador de determinar si un factor de riesgo se asocia a una respuesta porque los testigos han llegado a ser demasiado similares a los casos. Por ejemplo, como se emparejó a los testigos con los casos por la edad (± 3 años), el sexo, la raza y el nivel socioeconómico, el emparejamiento excesivo hizo que fuera imposible determinar si el nivel educativo se asociaba al riesgo de accidente cerebrovascular en los participantes de 65 años o mayores, porque las variables de emparejamiento son determinantes importantes del nivel educativo en ese grupo de edad.

Encuesta. Estudio transversal en una población específica, que habitualmente supone la administración de un cuestionario. Por ejemplo, en la encuesta *National Epidemiologic Survey on Alcohol and Related Conditions* se incluyó una muestra representativa de adultos estadounidenses y se les hicieron preguntas sobre el consumo de alcohol actual y pasado, los trastornos por consumo de alcohol y la utilización de servicios de tratamiento del alcoholismo.

Enmascaramiento. Proceso de garantizar que los pacientes y/o los investigadores desconozcan el grupo (es decir, de intervención o testigo) al que se asigna a los participantes, habitualmente en el contexto de un estudio aleatorizado. También se denomina ocultación,

especialmente en estudios oftalmológicos. Por ejemplo, mediante la utilización de comprimidos de placebo idénticos y el mantenimiento de la lista de las asignaciones de los participantes fuera del centro, tanto los participantes como los investigadores (incluyendo los auxiliares de investigación) desconocían a qué pacientes se había tratado con el medicamento activo.

Ensayo aleatorizado con enmascaramiento. Diseño de estudios en el que se asigna aleatoriamente a los participantes elegibles a los grupos del estudio con una probabilidad determinada, y los investigadores, los participantes y el resto del personal que participa en el estudio desconocen la asignación a los grupos del estudio. Por ejemplo, en un ensayo aleatorizado con enmascaramiento de un nuevo comprimido para el tratamiento de la diarrea haría falta que se asignara aleatoriamente a los participantes elegibles al nuevo comprimido o a un comprimido de placebo idéntico (habitualmente con una probabilidad del 50 % de ser asignado a cualquiera de los grupos), y que los investigadores, los participantes y el personal del estudio desconocieran si un participante está tomando el medicamento activo o el placebo.

Ensayo de ausencia de inferioridad. Ensayo clínico en el que se compara un nuevo tratamiento que tiene algunas ventajas respecto a un tratamiento establecido (p. ej., el nuevo tratamiento es más seguro, menos costoso o más fácil de utilizar), con el objetivo de demostrar que la eficacia del nuevo tratamiento no es inferior a la del tratamiento establecido. Por ejemplo, en un ensayo de un nuevo analgésico que no produce somnolencia, se demostró que el nuevo fármaco no era inferior a la oxicodona para el alivio del dolor postoperatorio.

Ensayo clínico. Diseño de investigación en el que los pacientes reciben una de (al menos) dos intervenciones diferentes. Habitualmente las intervenciones se asignan aleatoriamente; de aquí el término *ensayo clínico aleatorizado*. Los ensayos clínicos en ocasiones se denominan experimentos. Por ejemplo, el investigador realizó un ensayo clínico para determinar si el tratamiento profiláctico con penicilina reducía el riesgo de endocarditis bacteriana en pacientes con válvulas cardíacas anómalas a los que se realizaron intervenciones dentales.

Ensayo factorial. Ensayo clínico de dos o más tratamientos (p. ej., A y B), en ocasiones con dos variables de respuesta no relacionadas, en el que se asigna aleatoriamente a los pacientes a recibir un tratamiento activo A y un placebo de B, un tratamiento activo B y un placebo de A, los dos tratamientos activos A y B, o los dos placebos de A y B. Por ejemplo, el investigador realizó un estudio factorial para determinar si el uso a largo plazo de β-caroteno y ácido acetilsalicílico afectaba al riesgo de cáncer del tubo digestivo.

Ensayo de fase I. Ensayo de fase temprana, generalmente sin enmascaramiento y no controlado, en el que se evalúan dosis crecientes de un nuevo tratamiento en un número pequeño de voluntarios humanos para evaluar su seguridad. Por ejemplo, en un ensayo de fase I de un nuevo fármaco para el tratamiento de los sofocos menopáusicos generalmente se incluiría a un número pequeño de voluntarias (con o sin sofocos) a las que se administrarían dosis crecientes del fármaco para determinar sus efectos sobre el hemograma, la función hepática y renal, los hallazgos de la exploración física, los síntomas y otros acontecimientos adversos inesperados.

Ensayo de fase II. Pequeño estudio aleatorizado (y preferiblemente con enmascaramiento) para evaluar el efecto de diversas dosis de un nuevo tratamiento sobre los efectos adversos, además de sobre criterios de valoración indirectos o clínicos. Por ejemplo, en un ensayo de fase II de un nuevo tratamiento de los sofocos cuya seguridad se ha demostrado en un estudio de fase I, se podría incluir a un número pequeño de mujeres posmenopáusicas con sofocos, se las asignaría aleatoriamente a dos o tres dosis diferentes del nuevo fármaco o al placebo, y después se las seguiría para determinar la frecuencia de los sofocos, además de los efectos adversos.

Ensayo de fase III (fundamental). Ensayo aleatorizado (y preferiblemente con enmascaramiento) que es suficientemente extenso como para evaluar la eficacia y la seguridad de un nuevo tratamiento. Por ejemplo, si se ha establecido la dosis óptima de un nuevo tratamiento de los sofocos en un ensayo de fase II y el nuevo tratamiento era aceptablemente seguro, el siguiente paso sería un ensayo de fase III extenso en el que se asignaría aleatoriamente a mujeres posmenopáusicas con sofocos al nuevo tratamiento o al placebo, y se las seguiría para determinar la aparición de sofocos y de acontecimientos adversos.

Ensayo de fase IV. Extenso estudio, que puede ser o no un ensayo aleatorizado, realizado después de la autorización de comercialización de un fármaco por una agencia de registro sanitario como la U.S. Food and Drug Administration (FDA), habitualmente para determinar la seguridad del fármaco durante un período más prolongado de lo que es posible en un ensayo de fase III. Por ejemplo, después de la autorización por la FDA de un nuevo fármaco para el tratamiento de los sofocos menopáusicos, en un ensayo de fase IV se incluiría a mujeres con sofocos menos graves que aquellas a las que se incluyó en el ensayo de fase III.

Ensayo preclínico. Estudio que se realiza antes de evaluar una intervención en seres humanos. En estos ensayos se podrían incluir células, tejidos o animales. Por ejemplo, la U.S. Food and Drug Administration precisa ensayos preclínicos en dos especies de animales diferentes para documentar la seguridad antes de poder estudiar nuevos fármacos en seres humanos.

Epidemiología. Ciencia de determinar la frecuencia y los determinantes de las enfermedades o de otras variables de respuestas sanitarias en las poblaciones. Por ejemplo, en un estudio se investigó la epidemiología de la violencia por armas cortas en ciudades del interior.

Epidemiólogo. Médico, desglosado por sexo y edad. Por ejemplo, uno de los autores (¡pero no vamos a decir quién!).

Equilibrio. Situación en la que se desconoce cuál de dos posibilidades (el fármaco X es mejor que el placebo, el fármaco X es peor que el placebo) tiene más probabilidad de ser cierta. Por lo tanto, es ético comparar el fármaco X y el placebo en un ensayo aleatorizado. Por ejemplo, los investigadores pensaban que había un equilibrio clínico en un ensayo porque se desconocía si un nuevo tratamiento propuesto para el cáncer de esófago produciría mejores resultados que el tratamiento actual.

Error aleatorio. Separación de una medición o estimación respecto al valor verdadero debido a variación aleatoria. El error aleatorio se puede reducir repitiendo las mediciones y aumentando el tamaño de la muestra. Por ejemplo, si la prevalencia verdadera del consumo de aceite de pescado en personas con arteriopatía coronaria en la población es el 20 %, en un estudio en el que se incluya a 100 participantes se podría encontrar que exactamente el 20 % consume aceite de pescado, aunque simplemente por error aleatorio es probable que la proporción sea algo mayor o menor que esa cifra.

Error de clasificación. Error de medición de una variable categórica en la que se considera (erróneamente) que los participantes con un valor de la variable tienen otro valor. Por ejemplo, los autores estaban preocupados de que como las historias clínicas estaban incompletas, se considerara erróneamente que no se hubieran caído algunos participantes que realmente se hubieran caído durante el ingreso hospitalario. Véanse también *Error de clasificación diferencial* y *Error de clasificación no diferencial*.

Error de medición. Situación en la que la precisión o la exactitud (o ambas) de una medición son peor que perfectas; por lo tanto, siempre hay al menos algún error de medición para alguna de las variables (con la posible excepción de la muerte). Por ejemplo, para reducir el error de medición el investigador utilizó una pesa de acero inoxidable de 2 kg para calibrar semanalmente la balanza para lactantes.

Error sistemático. Véase *Sesgo*.

Error típico de la media. Estimación de la precisión de la media de una variable continua en una muestra; depende de la desviación típica (y de la raíz cuadrada) del tamaño de la muestra. Por ejemplo, el investigador señaló que la media de edad de la cohorte de 450 hombres era de 59 años, con un error típico de 0,48 años.

Error de tipo I. Error que se produce cuando se rechaza una hipótesis nula que realmente es verdadera en la población debido a un resultado estadísticamente significativo en un estudio. Por ejemplo, se produce error de tipo I si un estudio de los efectos del caroteno de la dieta sobre el riesgo de presentar cáncer de colon (con un valor de a de 0,05) concluye que el caroteno reduce el riesgo de cáncer de colon ($p < 0,05$) cuando realmente no hay ninguna asociación. Véase también *Resultado positivo falso*.

Error de tipo II. Error que se produce cuando el estudio no permite rechazar una hipótesis nula que realmente es falsa en la población (es decir, $p > \alpha$). Por ejemplo, se produce error de tipo II si un estudio no rechaza la hipótesis nula de que el caroteno no tiene ningún efecto sobre el riesgo de cáncer de colon ($p > 0,05$), cuando el caroteno realmente reduce el riesgo de cáncer de colon. Véase también *Resultado negativo falso*.

Escala. Abordaje frecuente para medir conceptos abstractos haciendo múltiples preguntas a las que se asigna una puntuación, y a las que se combina en una escala. Por ejemplo, la escala SF-36 para medir la calidad de vida hace 36 preguntas que permiten obtener ocho escalas relacionadas con la salud y el bienestar funcional. (SF es el acrónimo de *«short form»,* formato abreviado.) Véase también *Escala de Likert*.

Escala analógica visual. Escala (habitualmente una línea) que representa un espectro continuo de respuestas, desde un extremo hasta el otro. Habitualmente la línea mide 10 cm de longitud y la puntuación se mide como la distancia en centímetros desde el extremo inferior. Por ejemplo, una escala analógica visual de la intensidad del dolor podría estar formada por una línea recta con «sin dolor» en un extremo y «dolor insoportable» en el otro extremo; el participante en el estudio marca una «X» en el punto que mejor describe la intensidad del dolor que siente.

Escala de Likert. Conjunto de respuestas (habitualmente cinco) a una pregunta que ofrece un intervalo de opciones con un espacio similar entre ellas. Por ejemplo, las posibles respuestas a la pregunta «¿Cuál es la probabilidad de que vuelva usted a este servicio de urgencias para ser atendido?» fueron las siguientes: muy probable, algo probable, ni probable ni improbable, algo improbable, muy improbable.

Especificación. Estrategia de la fase de diseño para abordar un factor de confusión especificando un valor de esa variable de confusión como criterio de inclusión en el estudio. Por ejemplo, en un estudio del efecto del chupete sobre el riesgo de síndrome de muerte súbita del lactante, el investigador podría utilizar la especificación para incluir en el estudio únicamente a lactantes alimentados con fórmula. Si se encontrara una disminución del riesgo de muerte súbita en los niños que hayan utilizado el chupete, no se podría deber a que tuvieran más probabilidad de recibir lactancia materna.

Especificidad. Proporción de personas sin la enfermedad que se estudia en las que la prueba es negativa («negativa en la salud», o NS). Por ejemplo, en comparación con los resultados anatomopatológicos de la biopsia, la especificidad de un resultado > 4,0 ng/ml en el análisis del PSA sérico es de aproximadamente el 95 % para la detección de cáncer de próstata; en otras palabras, el 95 % de los hombres sin cáncer de próstata tendrán un PSA ≤ 4,0 ng/ml. Véase también *Sensibilidad*.

Estandarización. Instrucciones detalladas y específicas sobre cómo se debe realizar una medición, diseñadas para maximizar la reproducibilidad y la precisión de la misma. Por ejemplo, en un estudio en el que se mide la presión arterial, la estandarización de la medición podría incluir instrucciones sobre la preparación del participante, el tamaño del manguito que se debe utilizar, dónde se debe colocar el manguito, cómo inflar y desinflar el manguito, y qué ruidos cardíacos indican la presión arterial sistólica y diastólica.

Estratificación. Estrategia de la fase de análisis para controlar la confusión, separando a los participantes en el estudio en estratos de acuerdo con los niveles de una posible variable de confusión, y analizando la asociación entre las variables predictiva y de respuesta por separado en cada uno de los estratos. Por ejemplo, en un estudio de la asociación entre el ejercicio y el riesgo de accidente cerebrovascular, la falta de ejercicio habitual se podría asociar a aumento del riesgo de accidente cerebrovascular porque muchas personas que no hacen ejercicio son obesas, y la obesidad aumenta el riesgo de accidente cerebrovascular. Para reducir al mínimo el posible efecto de confusión de la obesidad, se estratificó a los participantes por el índice de masa corporal, y se realizaron los análisis por separado en las personas que tenían peso normal, sobrepeso u obesidad en situación inicial.

Estudio analítico. Estudio que busca asociaciones entre dos o más variables. Por ejemplo, el investigador realizó un estudio analítico de si la altura se correlacionaba con la presión arterial en estudiantes de medicina. Véase también *Estudio descriptivo*.

Estudio de antes/después. Estudio en el que se comparan los atributos de los participantes antes y después de una intervención. Por ejemplo, en el estudio se compararon las concentraciones séricas medias de colesterol antes y después del inicio de una dieta pobre en grasas.

Estudio de casos y cohortes. Diseño de investigación en el que se selecciona a los participantes que presentan una enfermedad (u otra variable de respuesta) como casos durante el seguimiento de una cohorte más extensa, y después se los compara con una muestra aleatoria de la cohorte total. Por ejemplo, en un estudio de casos y cohortes se incluyó una cohorte de 2 000 hombres con cáncer de próstata temprano, y se compararon las concentraciones de andrógenos y de vitamina D de muestras obtenidas en la situación inicial en los que murieron durante el seguimiento con las concentraciones de una muestra aleatoria de toda la cohorte.

Estudio de casos y testigos. Diseño de investigación en el que se comparan los casos que tienen una enfermedad (u otra variable de respuesta) con los testigos que no la tienen. Por ejemplo, en un estudio de casos y testigos se comparó el consumo semanal medio de frutos secos y semillas de los casos de diverticulitis vistos en un servicio de urgencias con el consumo de frutos secos y semillas de los testigos que tenían otros diagnósticos del tubo digestivo.

Estudio de casos y testigos anidado. Estudio de casos y testigos en el que se selecciona a los casos y los testigos de una cohorte definida (de mayor tamaño) o de entre los participantes incluidos previamente en un estudio de cohortes. Este estudio se utiliza generalmente cuando es demasiado costoso realizar determinadas mediciones en todos los participantes de la cohorte; por el contrario, se realizan esas mediciones en muestras de esos participantes que se habían almacenado en situación inicial. Por ejemplo, los investigadores realizaron un estudio de casos y testigos anidado para determinar si las concentraciones de citocinas en las muestras de sangre obtenidas en papel de filtro para el cribado neonatal se asociaban a la aparición de parálisis cerebral en la cohorte natal de 2009 del estado de Ohio.

Estudio de cohortes. Un estudio de cohortes prospectivo supone la inclusión de un grupo de participantes (la cohorte), la realización de algunas mediciones iniciales, y después su seguimiento a lo largo del tiempo para observar la variable de respuesta; en un estudio de cohortes retrospectivo se identifica un grupo de participantes (la cohorte) en los que ya se han realizado las mediciones y en los que ya se ha producido todo o parte del seguimiento. Por ejemplo, un investigador realizó un estudio de cohorte retrospectivo para determinar si el resultado de una prueba de inteligencia emocional aplicada cuando los soldados se enrolaron en el ejército estadounidense se asociaba a la probabilidad posterior de presentar trastorno de estrés postraumático (TEPT).

Estudio de cohortes dobles. Diseño de estudio en el que se incluye a los pacientes en una de dos cohortes distintas, a menudo por su ocupación. Por ejemplo, se utilizó un estudio de dobles cohortes para comparar los riesgos de dermatitis de contacto en las manos, además de las micosis de los pies, en alfareros y bailarines.

Estudio de cohortes múltiples. Estudio de cohortes en el que se incluyen dos o más grupos distintos de participantes (las cohortes), y después se comparan sus variables de respuesta. A menudo se utiliza en estudios de exposiciones laborales, en los que las cohortes que se comparan han estado expuestas a un posible factor de riesgo o no lo han estado. Por ejemplo, los investigadores realizaron un estudio de cohortes múltiples de si la exposición a los rayos cósmicos durante los vuelos en avión se asocia a un aumento del riesgo de neoplasias malignas sanguíneas; los investigadores estudiaron cuatro cohortes: pilotos y auxiliares de vuelo (que estarían expuestos a rayos

cósmicos) y vendedores de billetes y auxiliares de tierra (que no lo estarían). Véase también *Estudio de cohortes dobles*.

Estudio de cohortes prospectivo. Diseño de estudio en el que se miden los valores iniciales de las variables predictivas en un grupo definido (la cohorte) de participantes en el estudio, y después se sigue a ese grupo a loΣlargo del tiempo para detectar respuestas específicas. Por ejemplo, el Nurses Health Study es un estudio de cohortes prospectivo de factores de riesgo de enfermedades frecuentes en mujeres. La cohorte es una muestra de enfermeras de Estados Unidos, y las respuestas han incluido enfermedades cardiovasculares, cáncer y mortalidad.

Estudio de cohortes retrospectivo. Estudio de cohortes en el que la reunión de la cohorte, las mediciones iniciales y el seguimiento se han producido en el pasado. Por ejemplo, para describir la evolución natural de los aneurismas de la aorta torácica, un investigador que realizó un estudio de cohortes retrospectivo en 2012 podría obtener datos de registros del alta hospitalaria de pacientes a los que se realizó un diagnóstico de aneurisma aórtico en 2007, y utilizaría los registros del alta hospitalaria y el registro de mortalidad nacional para determinar en qué pacientes se produjo posteriormente rotura del aneurisma aórtico o la muerte antes de 2012.

Estudios cruzados. Variante del diseño de casos y testigos en el que cada caso sirve como su propio testigo, y se compara el valor de una exposición específica dependiente del tiempo en el período previo a que se produjera la variable de respuesta con su valor durante uno o más períodos de tiempo testigos. Este diseño está expuesto a sesgo de recuerdo, por lo que tiene su máxima utilidad cuando se puede determinar objetivamente una exposición. Por ejemplo, se utilizó un diseño de casos con cruzamiento para determinar si los pacientes que consultaron en un servicio de urgencias con migraña tenían más probabilidades de haber comido chocolate en las 2 h previas que a hora similar el día antes.

Estudio con cruzamiento. Diseño de investigación en el que se cambia a todos los participantes de un grupo de tratamiento (o testigo) al otro grupo, habitualmente en el punto medio del estudio. En ocasiones hay un período de reposo farmacológico entre las dos fases. Este diseño, que permite que todos los participantes reciban el tratamiento activo, sólo es útil para situaciones que vuelven al estado inicial después del tratamiento. Por ejemplo, se incluyó a pacientes con migraña en un estudio con cruzamiento para comparar un nuevo fármaco con el placebo para la prevención de las migrañas.

Estudio descriptivo. Estudio que no busca asociaciones, comprobación de hipótesis o realización de comparaciones. Por ejemplo, el investigador realizó un estudio descriptivo de la prevalencia de la obesidad en niños preescolares. Véase también *Estudio analítico*.

Estudio de equivalencia. Estudio cuya finalidad es demostrar que dos (o más) tratamientos producen resultados similares; habitualmente, uno de los tratamientos es nuevo y se sabe que el otro es eficaz. Por ejemplo, se utilizó un diseño de estudio de equivalencia para comparar dos antibióticos (fármaco nuevo A con fármaco antiguo B) para el tratamiento de la neumonía.

Estudio de hipótesis múltiples. Situación en la que un investigador analiza más de una (y habitualmente muchas más de una) hipótesis en un estudio, lo que aumenta el riesgo de cometer un error de tipo I salvo que se ajuste el nivel de significación estadística. Por ejemplo, aunque el investigador encontró una asociación estadísticamente significativa ($p = 0,03$) entre el consumo de vitamina D y el deterioro cognitivo, se criticaron sus resultados porque no había tenido en consideración el efecto del estudio de hipótesis múltiples, porque en el estudio se habían analizado más de 30 suplementos nutricionales. Véase también *Corrección de Bonferroni*.

Estudio de observación. Término general para un diseño de investigación en el que los investigadores simplemente observan a los participantes sin realizar ninguna intervención. Por lo tanto, esta categoría incluye estudios transversales, de casos y testigos y de cohortes, pero no ensayos aleatorizados ni estudios de antes y después. Por ejemplo, los investigadores realizaron un estudio de observación para determinar los factores de riesgo del melanoma.

Estudio piloto. Pequeño estudio realizado para determinar si un estudio a gran escala es viable, además de optimizar la logística y maximizar la eficiencia del estudio a gran escala. Por ejemplo, en un ensayo piloto de yoga restaurador para la prevención de la diabetes en pacientes con resistencia insulínica se pretendería demostrar la viabilidad de medir la resistencia insulínica, mejorar y normalizar la intervención con yoga, y demostrar que es posible incluir y aleatorizar a los participantes a los grupos de yoga y testigo.

Estudio de pruebas diagnósticas. Estudio en el que se evalúa si los resultados de una técnica médica son útiles para evaluar la probabilidad de un diagnóstico particular en un paciente. Por ejemplo, se elaboró un estudio de una prueba diagnóstica para determinar si la concentración sérica de bicarbonato es útil para diagnosticar sepsis en pacientes con fiebre.

Estudios de pruebas médicas. Término general utilizado para los estudios que miden en qué grado una prueba (o una serie de pruebas) identifica a los pacientes con un diagnóstico o una variable de respuesta particular. Por ejemplo, el investigador realizó un estudio de una prueba médica para determinar las razones de verosimilitud de la presencia y la ausencia de angina típica (definida como dolor torácico u opresión subesternal al esfuerzo) para el diagnóstico de arteriopatía coronaria.

Estudio de reproducibilidad. Estudio en el que la reproducibilidad de una medición es la principal pregunta de la investigación;

habitualmente se realiza comparando los resultados de una medición realizada muchas veces por la misma persona o con el mismo dispositivo (reproducibilidad intraobservador), o los resultados de la misma medición realizada por diferentes personas o con diferentes equipos (reproducibilidad interobservador). Por ejemplo, los investigadores realizaron un estudio de reproducibilidad para determinar si un nuevo estetoscopio electrónico podía mejorar la capacidad de detectar soplos cardíacos diastólicos.

Estudio transversal. Diseño de investigación en la que se selecciona a los participantes y se realizan las mediciones durante un período de tiempo limitado, habitualmente para estimar la prevalencia de una exposición o una enfermedad. Por ejemplo, se estimó la prevalencia de la miopía en un estudio transversal de 1 200 estudiantes universitarios de Berkeley, California.

Exactitud. El grado en el que una medición corresponde a su valor verdadero. Por ejemplo, el peso corporal autorreferido es una medición menos exacta del peso corporal real que el peso medido con una báscula electrónica calibrada.

Experimento. En investigación clínica, estudio en el que se asigna aleatoriamente a los participantes a un (o más) tratamiento o a un grupo de comparación. También se denomina ensayo aleatorizado. Por ejemplo, los investigadores realizaron un experimento para determinar si el fármaco X era mejor que el placebo en el tratamiento de la fibromialgia.

Exposición. Término utilizado para indicar que un participante en el estudio tiene un factor de riesgo particular. Por ejemplo, se definió la exposición al ácido acetilsalicílico como tomar un promedio de uno o más comprimidos de ácido acetilsalicílico (de cualquier dosis) a la semana durante el período previo de 6 meses.

Factor de confusión. Véase *Confusión*.

Heterogeneidad. Situación en la que la asociación entre una variable predictiva y otra de respuesta no es uniforme, ya sea entre estudios diferentes o entre distintos subgrupos de participantes. Por ejemplo, hay una heterogeneidad importante entre los estudios que han analizado los efectos del tratamiento posmenopáusico con estrógenos sobre el estado de ánimo y la cognición, de manera que en algunos estudios se han visto resultados positivos, en otros, efectos adversos, y en otros, ningún efecto.

Hipótesis. Término general para la declaración de una creencia sobre lo que se encontrará en el estudio. Por ejemplo, la hipótesis del estudio fue que el uso crónico de antiepilépticos se asociaba a un aumento del riesgo de cáncer bucal. Véanse también *Hipótesis nula* e *Hipótesis de la investigación*.

Hipótesis alternativa. Proposición, utilizada para estimar el tamaño de la muestra, de que hay una asociación entre las variables predictiva y de respuesta en la población. Por ejemplo, la hipótesis alternativa del estudio fue que los adolescentes que fuman cigarrillos tienen una probabilidad diferente de abandonar el colegio que aquellos que no fuman. Véase también *Hipótesis nula*.

Hipótesis bilateral. Hipótesis alternativa en la que el investigador está interesado en evaluar la posibilidad de cometer un error de tipo I en cualquiera de las dos direcciones posibles (es decir, riesgo mayor o menor). Por ejemplo, el investigador evaluó la hipótesis bilateral de que bailar salsa se asociaba a aumento o disminución del riesgo de demencia. Véase también *Hipótesis unilateral.*

Hipótesis compleja. Hipótesis de la investigación que tiene más de una variable predictiva o de respuesta. Deben evitarse las hipótesis complejas, porque es difícil verificarlas estadísticamente. Por ejemplo, los investigadores reformularon su hipótesis compleja («Que un nuevo programa de gestión de casos afectaría a la duración de la estancia y a la probabilidad de reingreso») en dos hipótesis simples («Que un nuevo programa de gestión de casos afectaría a la duración de la estancia» y también «Que un nuevo programa de gestión de casos afectaría a la probabilidad de reingreso»). Véase también *Hipótesis simple.*

Hipótesis de la investigación. Declaración del investigador que resume los elementos principales del estudio, como la población de interés, las variables predictivas y de respuesta, y el resultado previsto. Con fines estadísticos, la hipótesis de la investigación se plantea de una forma que establece la base de las pruebas de significación estadística, y generalmente incluye una hipótesis nula y una hipótesis alternativa. Por ejemplo, la hipótesis de la investigación era que la migraña se asociaría a un aumento del riesgo de accidente cerebrovascular de al menos el 20 %.

Hipótesis nula. Forma de la hipótesis de la investigación que especifica que no hay diferencia entre los grupos que se comparan. Por ejemplo, la hipótesis nula afirmaba que el riesgo de presentar claudicación sería el mismo en los participantes con concentraciones normales de lípidos a los que se trató con una estatina y en aquellos a los que se trató con el placebo.

Hipótesis***a posteriori.*** Hipótesis que se formulan después del análisis de los datos. Por ejemplo, en un estudio de la asociación entre el insomnio y el riesgo de accidente cerebrovascular, la hipótesis de que el insomnio aumenta el riesgo de diverticulitis es una hipótesis *a posteriori.*

Hipótesis simple. Hipótesis con una sola variable predictiva y una sola variable de respuesta. Por ejemplo, el investigador reformuló la

hipótesis compleja en la hipótesis simple de que las personas que consumen fruta al menos cinco veces a la semana tienen menos probabilidad de presentar cáncer de colon. Véase también *Hipótesis compleja*.

Hipótesis unilateral. Hipótesis alternativa en la que el investigador está interesado en evaluar la posibilidad de cometer un error de tipo I en tan sólo una de las dos direcciones posibles (es decir, riesgo mayor o menor, pero no ambos). Por ejemplo, el investigador estudió la hipótesis unilateral de que el tabaquismo se asociaba a aumento del riesgo de demencia. Véase también *Hipótesis bilateral*.

Homogeneidad. Situación en la que la asociación entre una variable predictiva y una variable de respuesta es uniforme en diferentes estudios. Por ejemplo, hay homogeneidad entre estudios de un tamaño razonable que han analizado los efectos del tabaquismo sobre el cáncer de pulmón: en todos se ha encontrado un gran aumento del riesgo en fumadores.

Incidencia. Proporción de personas que llegará a presentar una variable de respuesta durante el período de seguimiento; en ocasiones se denomina proporción de incidencia o incidencia acumulada. Por ejemplo, los investigadores encontraron que las embarazadas vegetarianas tenían menor incidencia de parto pretérmino que las embarazadas que comían carne.

Incidencia acumulada. Véase *Incidencia*.

Inclusión. Proceso de identificación y reunión de los participantes elegibles en un estudio. Los métodos de inclusión varían dependiendo de la naturaleza del estudio. Por ejemplo, la inclusión para el estudio supuso la identificación de pacientes elegibles en consultas especializadas, anuncios en folletos y periódicos, y utilización de Internet y páginas web de medios sociales.

Independiente. Este término se puede utilizar al menos de dos maneras. Primero, es la situación en la que las variables no influyen mutuamente unas sobre otras. Por ejemplo, los investigadores determinaron que el consumo de frutos secos con la dieta y la concentración sérica de glucosa eran independientes: no había datos en su estudio de que el consumo de frutos secos aumentara la concentración de glucosa, y viceversa. Segundo, se utiliza independiente para referirse a un efecto que tiene una variable sobre otra variable que no depende (es decir, «es independiente») de una tercera variable. Por ejemplo, como estaba preocupado de que el nivel educativo materno y la lactancia materna se asociaran entre sí, el investigador ajustó el nivel educativo materno para estimar el efecto independiente de la lactancia materna sobre las habilidades del lenguaje a los 2 años de edad.

Inferencia. Proceso de extraer conclusiones sobre una población basadas en las observaciones en una muestra. Por ejemplo, como refirió beber agua de pozo una cifra doble de casos de cáncer vesical que de testigos ($p = 0,02$), los investigadores realizaron la inferencia de que el consumo de agua de pozo aumenta el riesgo de cáncer vesical en la población.

Información sanitaria protegida. Información sanitaria identificable de manera individual. Las regulaciones federales sobre la privacidad de la salud (denominadas regulaciones HIPAA, por la *Health Insurance Portability and Accountability Act*) exigen que los investigadores mantengan la confidencialidad de la información sanitaria protegida en la investigación. Por ejemplo, la investigación sanitaria protegida no se debe almacenar en memorias portátiles ni se debe enviar por correo electrónico ordinario.

Interacción. Otro nombre para la *modificación del efecto*.

Intervalo de confianza. Término que habitualmente se interpreta mal; debe pensarse que un intervalo de confianza es una medida de la precisión: cuanto más estrecho sea el intervalo de confianza, más precisa es la estimación. Los intervalos de confianza están estrechamente relacionados con la significación estadística: un intervalo de confianza del (1 - a) % incluye (aproximadamente) el intervalo de valores cuya diferencia respecto a los valores observados no alcanzó la significación estadística (con un nivel de significación a). A menudo se interpreta erróneamente que los intervalos de confianza son afirmaciones directas sobre la *probabilidad posterior* (es decir, que hay una probabilidad del 95 % de que el valor verdadero esté contenido dentro del intervalo de confianza del 95 %). Esto es incorrecto, porque la probabilidad posterior depende de otra información, aparte de la que se encontró en el estudio. Por ejemplo, un riesgo relativo de 1,6 con un intervalo de confianza del 95 % desde 0,9 hasta 2,8 no sería estadísticamente significativo con un valor de α de 0,05, porque el intervalo incluye la «ausencia de efecto» (un riesgo relativo de 1,0). Véanse también *Alfa* y *Valor de* p.

Intervención. En un ensayo aleatorizado, el tratamiento activo que reciben los participantes. Muchas veces se utiliza como adjetivo (grupo de intervención). Por ejemplo, en un ensayo aleatorizado de psicoterapia para el tratamiento de la ansiedad, la intervención supuso 6 meses de sesiones semanales de 1 h con un psicólogo titulado e insistió en abordajes cognitivos-conductuales. Véase también *Testigo* (segunda definición).

Investigación aplicada. Investigación que tiene como objetivo traducir los hallazgos científicos en mejoras de la salud. La investigación aplicada puede tener como objetivo estudiar hallazgos de ciencia básica del laboratorio aplicándolos a estudios clínicos en pacientes (habitualmente denominados «investigación del laboratorio a la cabecera» o «T1»), o aplicar los hallazgos de los estudios clínicos para mejorar la salud de las poblaciones (a menudo denominados «investigación de la cabecera a la población» o «T2»). Por ejemplo, un estudio para determinar si una alteración genética que produce sordera congénita en ratones tiene un efecto similar en seres humanos

sería un estudio de investigación T1; un estudio para determinar si un proyecto estatal de cribado neonatal con una prueba que mide la respuesta cortical a los ruidos para detectar pérdida auditiva mejora el crecimiento escolar sería un estudio de investigación T2.

Investigador principal. Persona que tiene en último término la responsabilidad del diseño y la realización de un estudio, y del análisis y la presentación de los hallazgos del estudio. Por ejemplo, el comité ético de la investigación clínica pidió hablar con el investigador principal del estudio porque algunos miembros tenían preguntas sobre el protocolo.

Kappa (κ). Término estadístico que mide el grado en el que dos (o más) observadores están de acuerdo en si se ha producido o no un fenómeno, más allá de lo que cabe esperar por el azar. Varía desde -1 (desacuerdo perfecto) hasta 1 (acuerdo perfecto). Por ejemplo, el valor de κ del acuerdo entre dos anatomopatólogos sobre la presencia de cirrosis en una muestra de piezas de biopsia hepática fue de 0,85.

Magnitud del efecto. En relación con la planificación del tamaño de la muestra, la medida de cómo es de grande la diferencia que desea detectar el investigador entre los grupos que comparará, o la magnitud de la asociación. En un sentido más general, magnitud real de la diferencia o la asociación después de que haya finalizado el estudio. Por ejemplo, los investigadores basaron las estimaciones del tamaño de la muestra en una magnitud del efecto de una diferencia de 20 mg/dl de la glucemia entre los dos grupos.

Marcador indirecto. Medición que se piensa que se asocia a variables de respuesta clínica significativas. Un buen marcador indirecto habitualmente mide los cambios de un factor intermedio en la vía principal que determina la variable de respuesta clínica. Por ejemplo, el aumento del recuento de linfocitos CD4 en pacientes con infección por el virus de la inmunodeficiencia humana (VIH) es un buen marcador indirecto de la eficacia de los antirretrovíricos, porque predice un menor riesgo de infecciones oportunistas.

Marginales. Los totales de fila y columna de una tabla de contingencia. Por ejemplo, el estudio de los marginales de la tabla de 2 × 2 mostró que había números similares de hombres y mujeres en el estudio.

Media. Valor medio de una variable continua en una muestra o población; se calcula como la suma de todos los valores de la variable dividida por el número de participantes. Por ejemplo, la concentración media de colesterol sérico en una muestra de 287 mujeres de mediana edad era 223 mg/dl. Véanse también *Mediana* y *Desviación típica*.

Mediador. Variable que está producida por la variable predictiva de interés y que también da lugar a la variable de respuesta; explica al menos en parte *cómo* la variable predictiva produce la variable de respuesta. Por ejemplo, en un análisis del efecto de la obesidad sobre el riesgo de accidente cerebrovascular, los investigadores no controlaron la diabetes porque pensaban que un mecanismo mediante el cual la obesidad podría dar producir un accidente cerebrovascular era un mediador que producía diabetes.

Mediana. Valor de una variable que divide una muestra o población en dos mitades del mismo tamaño; es equivalente al percentil 50. A menudo se utiliza cuando una variable continua tiene algunos valores muy altos (o muy bajos) que influirían en exceso en la media. Por ejemplo, la mediana de los ingresos anuales de la muestra de 54 médicos fue de 225 000 dólares. Véanse también *Media* y *Desviación típica*.

Mediciones emparejadas. Mediciones muy relacionadas entre sí de alguna manera, como las que se realizan en lados diferentes de la misma persona, en miembros diferentes de una pareja de gemelos o (la mayoría de las veces) en el mismo participante en dos momentos diferentes, como antes y después de una intervención. Por ejemplo, en un estudio del efecto de un programa de ejercicio sobre la concentración de glucohemoglobina en participantes con diabetes del tipo II, las mediciones emparejadas de la glucohemoglobina incluyeron las mediciones realizadas en la situación inicial y de nuevo después de 3 meses de ejercicio.

Metaanálisis. Proceso para combinar los resultados de varios estudios con variables predictivas y de respuesta similares en un único resumen del resultado. Por ejemplo, en un metaanálisis de 12 estudios publicados, se encontró que el uso de antiinflamatorios no esteroideos se asociaba a un riesgo de presentar asma un 28 % mayor.

Método de referencia. Método no ambiguo para determinar si un paciente tiene o no una enfermedad o una variable de respuesta particular. Por ejemplo, el método de referencia para el diagnóstico de fractura de cadera precisaba la confirmación radiológica por un radiólogo certificado.

Modelo de Cox. También llamado *modelo de riesgos proporcionales de Cox*. Técnica estadística multivariada que mide los efectos individuales de una o más variables predictivas sobre la incidencia (riesgo) con la que se produce una variable de respuesta en la muestra, teniendo en consideración las diferentes duraciones del seguimiento en los distintos pacientes. Por ejemplo, utilizando un modelo de riesgos proporcionales de Cox, los hombres tenían aproximadamente el doble de probabilidad que las mujeres, y los negros aproximadamente el triple de probabilidad que los blancos, de presentar accidente cerebrovascular después de ajustar la edad, la presión arterial y la diabetes, además de la duración del seguimiento. Véase también *Modelo de regresión logística*.

Modelo de efectos aleatorios. Término utilizado en el análisis estadístico multinivel. Se describe en este libro únicamente en relación con el metaanálisis, donde describe un modelo estadístico en el que los pesos de los estudios y la varianza de la estimación resumen del efecto incorporan un término correspondiente a la variabilidad de los resultados de los estudios individuales incluidos. Por ejemplo, en un metaanálisis de ensayos clínicos del efecto de la práctica del yoga sobre la depresión, los resultados en los ensayos fueron variables; por lo tanto, los estudios pequeños contribuyeron más al resumen del efecto, y el intervalo de confianza fue más amplio utilizando el modelo de efectos aleatorios que con el modelo de efectos fijos. Véase también *Modelo de efectos fijos*.

Modelo de efectos fijos. Término general que se utiliza en análisis estadístico multinivel; en este libro sólo se utiliza en relación con el metaanálisis, donde describe un modelo estadístico en el que los pesos del estudio y la varianza de la estimación resumen del efecto se basan únicamente en las variancias intraestudio de los estudios incluidos. Por ejemplo, en un metaanálisis de ensayos clínicos del efecto de la práctica del yoga sobre la depresión, los resultados de los ensayos fueron variables; el resumen del efecto basado en el modelo de efectos fijos estuvo dominado por un estudio extenso, y el intervalo de confianza fue más estrecho de lo que se habría estimado con un modelo de efectos aleatorios. Véase también *Modelo de efectos aleatorios*.

Modelo de regresión logística. Técnica estadística que se utiliza para estimar los efectos de una o más variables predictivas sobre una variable de respuesta dicotómica, ajustando los efectos de otras variables predictivas y de confusión. Por ejemplo, en un modelo de regresión logística, los hombres tenían aproximadamente el doble de probabilidad que las mujeres, y los negros aproximadamente tres veces la probabilidad de los blancos, de presentar un accidente cerebrovascular, después de ajustar la edad, la presión arterial y la diabetes.

Modificación del efecto. Situación en la que la fuerza de la asociación entre una variable predictiva y una variable de respuesta se ve modificada por una tercera variable (denominada a menudo modificador del efecto, aunque puede ser difícil determinar cuál es la variable predictiva y cuál el modificador del efecto). Por ejemplo, los investigadores encontraron que los efectos del nivel de ingresos sobre el riesgo de accidente cerebrovascular diferían en blancos y negros, de manera que la pobreza tenía una mayor asociación con el accidente cerebrovascular en negros que en blancos. Véase también *Confusión*.

Muestra. Subgrupo de la población que participa en un estudio. Por ejemplo, en un estudio de un nuevo tratamiento del asma, en el que la población objetivo son todos los niños con asma y la población accesible son los niños con asma de la ciudad del investigador vistos este año, la muestra del estudio son los niños de la ciudad del investigador vistos este año que realmente participan en el estudio.

Muestra aleatoria. Muestra extraída mediante la enumeración de las unidades de la población y la selección de un subgrupo de manera aleatoria. Por ejemplo, la obtención de una muestra aleatoria de personas con cataratas de la consulta de un investigador precisaría que el investigador elaborara una lista de todos los pacientes con cataratas y que utilizara una tabla de números aleatorios o números aleatorios generados por ordenador para seleccionar la muestra. Véase también *Muestreo probabilístico*.

Muestra basada en la población. Muestra de personas que representan a una población completa. Por ejemplo, en el estudio National Health and Nutrition Examination Survey (NHANES), que aporta datos sobre una muestra aleatoria de toda la población de Estados Unidos, se utiliza una muestra basada en la población.

Muestra consecutiva. Muestra de un estudio en la que se elige a los pacientes uno tras otro hasta que se alcance el tamaño de la muestra. Habitualmente se utiliza para referirse a la muestra prevista; también se puede referir a la muestra real cuando se realizan revisiones de historias clínicas, porque puede no ser necesario el consentimiento informado. Por ejemplo, los investigadores realizaron muestreo consecutivo para revisar las historias de los primeros 100 pacientes con artritis reumatoide vistos en la consulta de reumatología desde el 15 de enero de 2013.

Muestra de conveniencia. Grupo de participantes a los que se seleccionó para un estudio simplemente porque era relativamente fácil acceder a ellos. Por ejemplo, el investigador utilizó una muestra de conveniencia de los pacientes de su consulta para que actuaran como testigos en su estudio de casos y testigos de los factores de riesgo de meningioma.

Muestra prevista. Grupo de pacientes que el investigador quiso incluir en un estudio, tal y como se describió en el protocolo del estudio. Por ejemplo, la muestra prevista para el estudio incluyó a mujeres con cáncer de mama a las que se vio inicialmente para su tratamiento los lunes o los jueves en el hospital Longview (los días que el investigador o su equipo de investigación estaban disponibles) y en las que habían transcurrido menos de 6 semanas desde el diagnóstico inicial, durante el período desde el 1 de enero de 2013 hasta el 30 de junio de 2014. Véanse también *Población accesible* y *Muestra*.

Muestra representativa. Muestra de personas incluidas en un estudio que representa a la población objetivo. Por ejemplo, en el *Framingham Heart Study,* la población objetivo eran todos los adultos. La población accesible (para los investigadores localizados en Boston) fue la población adulta de la ciudad de Framingham, Massachusetts. Los investigadores realizaron una lista de los adultos de Framingham y pidieron a uno de cada dos residentes que participaran en el estudio. Este abordaje podría llevar a una muestra

representativa, aunque algunas personas se negaron a participar y fueron sustituidas por voluntarios. Como los voluntarios generalmente tienen hábitos más saludables que los no voluntarios, la muestra puede haber tenido una representación excesiva de personas sanas. Además, la población de Framingham (que era principalmente blanca) no representa a todos los adultos de Estados Unidos, y es evidente que no representa a adultos de otros países.

Muestra sistemática. Muestra que se extrae enumerando las unidades de la población elegible y seleccionando un subgrupo de la población mediante un proceso preespecificado. Por ejemplo, en el *Framingham Heart Study,* los investigadores elaboraron una lista de todos los residentes adultos de la ciudad de Framingham, Massachusetts, y después seleccionaron a uno de cada dos residentes para su inclusión en el estudio como parte de una muestra sistemática.

Muestreo. Proceso de selección de participantes para su inclusión en un estudio cuando el número de participantes elegibles es mayor que el tamaño de la muestra estimado. Por ejemplo, el investigador utilizó un esquema de muestreo de «1 de 3» para seleccionar, en promedio, a uno de cada tres participantes elegibles. Véanse también *Muestreo por conglomerados, Muestreo consecutivo, Muestreo de conveniencia, Muestreo probabilístico, Muestreo aleatorio estratificado* y *Muestreo sistemático*.

Muestreo aleatorio estratificado. Técnica de muestreo en la que se estratifica a los posibles participantes en grupos de acuerdo con características como la edad, la raza o el sexo, y se obtiene una muestra aleatoria de cada uno de los estratos. Se pueden ponderar los estratos de diferentes maneras. Por ejemplo, los investigadores utilizaron muestreo aleatorio estratificado en un estudio de la prevalencia del cáncer pancreático en California para realizar un muestreo excesivo de las minorías raciales y étnicas.

Muestreo por conglomerados. Técnica de muestreo en la que se selecciona a los participantes como grupos (conglomerado) y no como individuos. Se utiliza la mayoría de las veces por comodidad cuando se realizan muestreos de poblaciones extensas. Por ejemplo, un investigador interesado en determinar la prevalencia del consumo de drogas utilizó muestreo por conglomerados para incluir a 300 pacientes. Primero identificó a los posibles participantes eligiendo 10 prefijos de tres cifras (p. ej., 285-, 336-, etc.) en un código de área, y después utilizó marcación telefónica aleatoria para encontrar 30 participantes que se prestaran voluntarios en cada uno de los conglomerados de tres dígitos.

Muestreo con densidad de incidencia. En un estudio de casos y testigos anidado, técnica para muestrear los testigos cuando una exposición importante cambia a lo largo del tiempo; por lo tanto, se debe medir la exposición en un momento similar en los casos y en los testigos. Por ejemplo, en un estudio de casos y testigos anidado para determinar si el uso de antihistamínicos, que varía estacionalmente, aumenta el riesgo a corto plazo de fracturas de cadera (probablemente debido a un aumento del riesgo de caída), se utilizó muestreo por densidad de incidencia de los testigos, de manera que el uso por un testigo del antihistamínico se midió el mismo mes en que se produjo una fractura de cadera en un caso.

Muestreo probabilístico. Proceso aleatorio, habitualmente utilizando una tabla de números aleatorios o un algoritmo informático, para garantizar que cada uno de los miembros de una población tenga una probabilidad especificada de ser incluido en la muestra, de manera que constituye una base rigurosa para hacer inferencias desde la muestra hasta la población. Por ejemplo, la observación de una muestra probabilística del 5 % de las personas con enfermedad pulmonar obstructiva crónica (EPOC) basada en los diagnósticos al alta hospitalaria de todos los hospitales de California debería ofrecer hallazgos fiables sobre los factores de riesgo de reingreso y muerte.

Nivel de significación estadística. Véase *Alfa*.

Normalización. En una base de datos relacional, proceso de eliminar las redundancias y mejorar la fiabilidad, asegurándose de que los datos de cada una de las preguntas se almacenen en no más filas o tablas de lo necesario. Por ejemplo, después de que el consultor de la base de datos hubiera normalizado la base de datos, pudo actualizar el número de teléfono de un participante alterando tan sólo una fila de una única tabla.

Número que es necesario tratar. Número absoluto de personas que deben recibir un tratamiento para prevenir la aparición de una variable de respuesta. Se calcula como el recíproco de la diferencia de riesgos. Por ejemplo, cuando se evaluaron los efectos beneficiosos del tratamiento de la hipertensión leve a moderada, el número que era necesario tratar era de aproximadamente 800 pacientes al año para prevenir un episodio de accidente cerebrovascular.

Objetivos específicos. En una propuesta de investigación, declaraciones breves sobre los objetivos de la investigación. Por ejemplo, un objetivo específico de un ensayo aleatorizado del efecto de la testosterona sobre la densidad mineral ósea en hombres podría ser: «Evaluar la hipótesis de que, en comparación con los hombres a los que se asigne a recibir el parche de placebo, aquellos a los que se asigne a recibir el parche de testosterona tendrán menos osteopenia durante 3 años de tratamiento».

Ocultación. Véase *Enmascaramiento*.

Partición recursiva. Técnica multivariada para clasificar a las personas según su riesgo de una variable de respuesta; al contrario de las

técnicas que precisan un modelo, como la regresión logística, no precisa ningún supuesto sobre la forma de la relación entre las variables predictivas y la variable de respuesta. Crea un árbol de clasificación que incluye una serie de preguntas con respuesta de sí/no, denominado árbol de clasificación y regresión (ACR). Por ejemplo, utilizando partición recursiva, los investigadores determinaron que los pacientes del servicio de urgencias de 20 a 65 años que tenían dolor abdominal pero que no tenían pérdida de apetito, fiebre ni dolor de rebote a la palpación, tenían riesgo bajo de apendicitis aguda. Véanse también *Regla de predicción clínica* y *Ajuste excesivo*.

Participante. Alguien que participa en un estudio de investigación. A menudo se prefiere el término *participante* a *sujeto,* porque insiste en que la persona incluida en el estudio es un participante activo en el avance de la ciencia, no simplemente un sujeto en el que se experimenta. Por ejemplo, en un estudio de un nuevo tratamiento del insomnio, los participantes son las personas que son elegibles y a las que se incluye en el estudio.

Período de preinclusión. En un ensayo clínico, período breve durante el cual los pacientes elegibles toman el placebo o se les aplica la intervención activa; sólo aquellos que consiguen un determinado nivel de cumplimiento, toleran la intervención o tienen un efecto beneficioso en una variable de respuesta intermedia son elegibles para el ensayo principal. Por ejemplo, en el estudio *Cardiac Arrhythmia Suppression Trial* sólo se aleatorizó a continuar la medicación o a cambiar al placebo a aquellos que tuvieron una reducción satisfactoria de las extrasístoles ventriculares con la medicación activa durante el período de preinclusión.

Período de reposo farmacológico. En un estudio con cruzamiento, tiempo transcurrido entre el primer y el segundo tratamiento, cuya finalidad es permitir que los efectos de la intervención se desvanezcan y que la medida de la variable de respuesta vuelva a la situación inicial. Por ejemplo, en un estudio cruzado en el que se compare un diurético con el placebo para el tratamiento de la hipertensión arterial, el investigador podría dejar un período de reposo farmacológico sin tratamiento entre los dos períodos de tratamiento para permitir que la presión arterial volviera al valor inicial.

Persona-tiempo. Suma de los períodos de tiempo durante los cuales cada uno de los participantes en un estudio o una población están en riesgo; se utiliza como denominador para el cálculo de las tasas de incidencia. Se puede calcular como el número de personas que tienen riesgo de una variable de respuesta multiplicado por el tiempo medio que están en riesgo. Por ejemplo, el número de persona-tiempo total de seguimiento en 1 000 participantes que tuvieron un promedio de 2,5 años de riesgo fue un total de 2 500 personas-año, aunque se siguió al 5 % de los participantes durante 1 mes o menos. Véase también *Tasa de incidencia*.

Personas vulnerables. Posibles participantes en el estudio que tienen mayor riesgo de ser utilizados de maneras éticamente inadecuadas en la investigación. Por ejemplo, como las personas con deterioro cognitivo o problemas de comunicación pueden no poder dar un consentimiento totalmente informado a participar en un estudio, se considera que son personas vulnerables. Otros ejemplos son niños, prisioneros, fetos y personas con un nivel socioeconómico bajo.

Plagio. Tipo de conducta científica impropia en la que el investigador se apropia de las ideas, resultados o palabras de otra persona sin otorgarle el crédito correspondiente. Por ejemplo, el uso de la descripción por otro investigador de un nuevo método de medición sin una atribución correcta constituye plagio.

Población. Conjunto completo de personas con las características especificadas. Por ejemplo, se podría definir la población adulta de Estados Unidos con diabetes del tipo II como todos los adultos estadounidenses que toman un fármaco hipoglucemiante o que tienen una glucemia basal mayor de 125 mg/dl.

Población accesible. Grupo de personas a las cuales el investigador tiene acceso y a las que se podría seleccionar, o abordar, para participar en el estudio. Por ejemplo, la población accesible para el estudio estaba formada por mujeres con cáncer de mama a las que se trató en las 6 semanas siguientes al diagnóstico original en el hospital Longview desde el 1 de enero de 2013 hasta el 30 de junio de 2014. Véanse también *Muestra prevista* y *Población objetivo*.

Población objetivo. Gran conjunto de personas definidas por características clínicas y demográficas a las que el investigador del estudio quiere generalizar los resultados del estudio. Por ejemplo, la población objetivo para un estudio de un nuevo tratamiento del asma en niños del hospital del investigador podrían ser los niños con asma de todo el mundo.

Posibilidad de generalización. Grado en el que se piensa que los resultados de una muestra de estudio se aplican a otras poblaciones. Por ejemplo, el revisor puso en duda la posibilidad de generalizar una tasa de éxito publicada del 90 % de la radioablación intraluminal de las redes esofágicas inferiores, porque todos los procedimientos los realizó el digestólogo que había inventado y perfeccionando la técnica en 350 pacientes, mientras que la mayoría de los digestólogos en ejercicio verían tan sólo un pequeño grupo de pacientes con el mismo problema a lo largo de su carrera profesional.

Posibilidades. Riesgo de una enfermedad (o de otra variable de respuesta) dividido por (1 - riesgo). Por ejemplo, si el riesgo de cáncer de mama a lo largo de toda la vida en mujeres es del 15 %, entonces la posibilidad de presentar cáncer de mama a lo largo de toda la vida

es 0,18 (0,15/0,85). El riesgo y la posibilidad son similares para enfermedades infrecuentes (las que se producen en menos de aproximadamente el 10 % de las personas).

Potencia. Probabilidad de rechazar correctamente la hipótesis nula en una muestra si el efecto real en la población es igual o mayor que una magnitud del efecto especificada. Por ejemplo, suponga que el ejercicio produce una reducción media de 20 mg/dl de la glucosa basal en mujeres diabéticas de toda la población. Si un investigador ajusta la potencia en el 90 % y extrae una muestra de la población en numerosas ocasiones, y realiza en todas las ocasiones el mismo estudio con las mismas mediciones, entonces en nueve de cada 10 estudios el investigador rechazaría correctamente la hipótesis nula y concluiría que el ejercicio reduce la glucemia basal. Véase también *Beta*.

Precisión. Grado en el que una medición de una variable es reproducible, con un valor que es casi el mismo en todas las ocasiones en que se mide. Por ejemplo, una báscula mecánica de balancín puede medir el peso corporal con mucha precisión, mientras que una entrevista para medir la gravedad de la depresión tiene más probabilidad de producir valores que varían de un observador a otro.

Pregunta de la investigación. La pregunta que pretende responder un proyecto de investigación. Una buena pregunta de la investigación debe incluir las variables predictiva y de respuesta de interés, y la población que se estudiará. Las preguntas de la investigación generalmente adoptan la forma de «¿Está asociado A con B en la población C?», o (en un ensayo clínico) «¿A produce B en la población C?» Por ejemplo, «¿El uso habitual de seda dental reduce el riesgo de episodios coronarios en pacientes diabéticos?»

Pregunta de la investigación secundaria. Preguntas distintas a la pregunta de la investigación principal, que a menudo incluyen factores predictivos o variables de respuesta adicionales. Por ejemplo, si la pregunta de la investigación principal es determinar la asociación entre el consumo de alcohol en embarazadas y el bajo peso al nacimiento en los lactantes, una pregunta secundaria sería determinar la asociación entre el consumo de alcohol y la anemia durante el embarazo.

Prevalencia. Proporción de personas que tienen una enfermedad o trastorno en un momento determinado. La prevalencia depende tanto de la incidencia de una enfermedad como de su duración. Por ejemplo, la prevalencia del lupus eritematoso sistémico es la proporción de personas que tienen esta enfermedad en un momento determinado; podría aumentar si la enfermedad fuera más frecuente o si el tratamiento produjera tal mejoría que las personas con la enfermedad vivieran más tiempo.

Propuesta. Documento que incluye un protocolo del estudio, un presupuesto y otra información administrativa y de apoyo, que se elabora para obtener financiación de una agencia financiadora. Por ejemplo, los National Institutes of Health (NIH) precisan propuestas para la financiación de múltiples tipos de investigación.

Propuesta de investigación. Documento elaborado para obtener financiación para la investigación, en el que se describe el diseño del estudio propuesto, los participantes, las mediciones, los análisis estadísticos y los aspectos éticos. Por ejemplo, los National Institutes of Health reciben miles de propuestas de investigación cada año de investigadores que solicitan financiación para sus estudios.

Protocolo. Plan de un estudio escrito y detallado. Por ejemplo, en el protocolo de un estudio se especificó que solamente se podía elegir como participantes a las personas que entendieran el inglés con un nivel de octavo curso.

Prueba de la ji al cuadrado (χ^2). Técnica estadística que compara dos (o más) proporciones para determinar si difieren significativamente entre sí. Por ejemplo, en un estudio se determinó si el riesgo de demencia era similar en personas que realizaban ejercicio al menos dos veces a la semana y en personas que realizaban ejercicio con menos frecuencia, comparando estadísticamente esos riesgos con una prueba de la χ^2.

Prueba preliminar. Evaluación de cuestionarios, medidas o procedimientos específicos que puede realizar el personal del estudio antes del comienzo de un estudio. Su finalidad es evaluar la funcionalidad, adecuación o viabilidad de la medición. Por ejemplo, la prueba preliminar del sistema de introducción de datos y gestión de bases de datos se podría realizar pidiendo al personal del estudio que rellenara formularios con datos ausentes, fuera del intervalo o ilógicos para garantizar que el sistema de edición de los datos identificara esos errores.

Prueba de la *t* (o prueba de la *t* de Student). Prueba estadística que se utiliza para determinar si el valor medio de una variable continua en un grupo difiere significativamente del de otro grupo. Por ejemplo, en los participantes en un estudio a los que se trató con dos antidepresivos diferentes podría utilizarse la prueba de la *t* para comparar las puntuaciones medias de depresión después del tratamiento de los dos grupos (una prueba de la *t* para dos muestras no emparejadas) o el cambio medio desde el valor inicial hasta después del tratamiento en los dos grupos (prueba de la *t* para dos muestras emparejadas). Véase también *Prueba de la* t *para una muestra* y *Prueba de la* t *para dos muestras.*

Prueba de la *t* para una muestra. Prueba estadística que se utiliza para comparar el valor medio de una variable en una muestra con una

constante fija (un número particular). El tipo más frecuente de prueba de la *t* de una muestra es la *prueba de la* t *emparejada,* en la que la media de la diferencia entre mediciones emparejadas de una muestra (p. ej., en el mismo participante en diferentes momentos) se compara con cero. Por ejemplo, los investigadores encontraron que los hombres ganaban una media (± DT) de 4 ± 3 kg de peso durante el período de residencia (p = 0,03 mediante la prueba de la *t* para una muestra). Véase también *Prueba de la* t *para dos muestras*.

Prueba de la***t*** para dos muestras. Prueba estadística que se utiliza para comparar el valor medio de una variable en una muestra con su valor medio en otra muestra. Por ejemplo, los investigadores encontraron que los participantes tratados con suplementos de aceite de oliva tuvieron un aumento medio de la concentración de colesterol unido a lipoproteínas de alta densidad de 10 mg/dl durante el estudio, en comparación con un aumento de 2 mg/dl en los que recibieron el placebo (p = 0,14 mediante la prueba de la *t* para dos muestras). Véase también *Prueba de la* t *para una muestra*.

Prueba de la***Z.*** Prueba estadística utilizada para comparar proporciones a fin de determinar si hay diferencias estadísticamente significativas entre ellas. Al contrario de la prueba de la χ^2, que siempre es bilateral, se puede utilizar la prueba de la *Z* para hipótesis unilaterales. Por ejemplo, puede utilizarse una prueba de la *Z* unilateral para determinar si la proporción de presos que tienen diabetes es significativamente mayor que la proporción de personas que viven fuera de la cárcel que tienen diabetes. De manera similar, se podría utilizar una prueba de la *Z* (o una prueba de la χ^2) bilateral para determinar si la proporción de presos que tienen diabetes es significativamente diferente (es decir, mayor o menor) de la proporción de personas de fuera de la prisión que tienen diabetes.

Puntuación de susceptibilidad. Probabilidad estimada de que un participante en un estudio tenga un valor especificado de una variable predictiva, la mayoría de las veces la probabilidad de recibir un tratamiento particular. El control de la puntuación de susceptibilidad (p. ej., mediante emparejamiento, estratificación o análisis multivariado) es un método para afrontar la confusión por indicación: en lugar de ajustar todos los factores que se podrían asociar a la variable de respuesta, el investigador crea un modelo multivariado para predecir la recepción del tratamiento. Después se asigna a cada uno de los pacientes una probabilidad prevista de tratamiento (la puntuación de susceptibilidad), que después se puede utilizar como única variable de confusión cuando se estime la asociación entre el tratamiento y la variable de respuesta. Por ejemplo, los investigadores utilizaron una puntuación de susceptibilidad para ajustar los valores asociados al uso de ácido acetilsalicílico a fin de determinar la asociación entre el uso de ácido acetilsalicílico y el cáncer de colon.

Razón de posibilidades. Cociente de las posibilidades de una enfermedad (o de otra variable de respuesta) en las personas expuestas a un factor de riesgo entre las posibilidades de esa enfermedad en las personas no expuestas. El cociente de riesgos y la razón de posibilidades son similares cuando una enfermedad es infrecuente en personas expuestas y no expuestas, porque las posibilidades y los riesgos de la enfermedad son similares. Por ejemplo, la razón de posibilidades de insuficiencia renal en personas con hipertensión es de 2,0, lo que significa que los pacientes hipertensos tienen aproximadamente el doble de probabilidad de presentar insuficiencia renal que las personas no hipertensas.

Razón de riesgos (riesgo relativo). Riesgo de una variable de respuesta en un grupo dividido por el riesgo en un grupo de comparación. Por ejemplo, si el riesgo de episodios tromboembólicos venosos en mujeres que utilizan actualmente estrógenos es 5/1 000 (0,5 %) y el riesgo en las que nunca han utilizado estrógenos es 2/1 000 (0,2 %), el riesgo relativo de las mujeres que utilizan estrógenos en comparación con las que no lo hacen es 2,5. Véanse también *Cociente de riesgos instantáneos* y *Razón de posibilidades*.

Razón de verosimilitud. Término que se utiliza para describir los efectos cuantitativos del resultado de una prueba médica sobre la probabilidad de que un paciente tenga la enfermedad que se estudia. Se define como la probabilidad de ese resultado de la prueba en un paciente *con* la enfermedad dividido por la probabilidad de ese resultado en un paciente *sin* la enfermedad. Por ejemplo, el cociente de verosimilitudes de los síntomas característicos de la angina típica (opresión subesternal con el esfuerzo) es de aproximadamente 50 para el diagnóstico de arteriopatía coronaria.

Redes de investigación basadas en consultas. Redes en las que los médicos de contextos comunitarios trabajan juntos para estudiar preguntas de la investigación de interés. Por ejemplo, en un estudio de una red de investigación basada en consultas del tratamiento del síndrome del túnel carpiano en atención primaria, se observó que la mayoría de los pacientes mejoraban con tratamiento conservador. Esto contrastaba con la bibliografía previa de centros médicos académicos, en la que se observaba que la mayoría de los pacientes con síndrome del túnel carpiano necesitaban cirugía.

Registro***(record).*** Fila de una tabla de bases de datos relacional (que se identifica de manera óptima por una *clave principal*) que incluye información sobre esa persona, transacción, resultado o fenómeno. Por ejemplo, una tabla de participantes podría tener un registro para cada uno de los participantes del estudio, con IdEstudio como clave principal, además de otra información, como la fecha de nacimiento y el sexo como campos.

Registro***(registry).*** Base de datos de personas con una enfermedad determinada o a las que se ha realizado una intervención particular. Los estudios se pueden realizar utilizando registros y recogiendo los datos de las variables de respuesta como parte del registro, o

relacionando los datos de registros con otras fuentes, como registros de cáncer o el registro de mortalidad nacional. Por ejemplo, en el San Francisco Mammography Registry se obtienen datos sobre todas las mujeres a las que se realiza una mamografía en los tres principales centros de mamografía de San Francisco; los investigadores han relacionado estos datos con registros de cáncer locales para estimar la exactitud de la mamografía.

Regla de predicción clínica. Algoritmo que combina diferentes variables predictivas, como la presencia o ausencia de diversos síntomas o signos y los resultados de las pruebas médicas, para estimar la probabilidad de una enfermedad o una variable de respuesta particular. Por ejemplo, los investigadores elaboraron una regla de predicción clínica para el diagnóstico de fractura de muñeca en mujeres posmenopáusicas basada en la información sobre fracturas previas, las características de la caída (si se había producido), la exploración física del antebrazo y los medicamentos actuales.

Regresión a la media. Tendencia a que los valores extremos (muy altos y muy bajos) sean más próximos a la media poblacional cuando se repiten. Por ejemplo, en un grupo de niños seleccionados para un estudio por tener presiones arteriales sistólicas por encima del percentil 95, se observó que la mayoría de los niños tenían menores valores de presión arterial en la primera visita de seguimiento, aunque todavía no habían recibido ningún tratamiento.

Respuesta. Término general para los criterios de valoración de un estudio, como la muerte o la aparición de una enfermedad. Por ejemplo, en un estudio de si la radiocirugía era beneficiosa en pacientes con metástasis cerebrales solitarias, se siguió a los pacientes para determinar la respuesta de muerte o internamiento en una residencia de ancianos especializada.

Resultado negativo falso. Término que se puede utilizar de dos maneras distintas. En el contexto de una prueba médica, se refiere a un resultado de una prueba que es falsamente negativo en un paciente que tiene la enfermedad que se estudia. Por ejemplo, aunque la paciente tenía cáncer de mama confirmado mediante biopsia, la mamografía dio un resultado negativo falso. En el contexto de un estudio de investigación se refiere a un estudio que no detecta en la muestra un efecto (es decir, el resultado del estudio no es estadísticamente significativo) que sí está presente en la población. Por ejemplo, aunque en estudios posteriores se observó que el tabaquismo aumenta el riesgo de accidente cerebrovascular, en uno de los primeros estudio de casos y testigos se observó un resultado negativo falso (p = 0,23).

Resultado positivo falso. Término que se puede utilizar de dos maneras diferentes. En el contexto de una prueba médica se refiere a un resultado de una prueba que es falsamente positivo en un paciente que no tiene la enfermedad que se estudia. Por ejemplo, aunque la paciente no tenía cáncer de mama ni lo presentó durante 6 años de seguimiento, la mamografía mostró un resultado falsamente positivo. En el contexto de un estudio de investigación, se refiere a un estudio que detecta en la muestra un efecto (es decir, el resultado del estudio es estadísticamente significativo) que no está presente en la población. Por ejemplo, aunque en estudios posteriores se demostró que el tabaquismo no aumenta el riesgo de enfermedad de Parkinson, se observó un resultado positivo falso en uno de los primeros estudios de casos y testigos (p = 0,03).

Resumen del efecto. En un metaanálisis, el efecto medio ponderado que se observa en los estudios incluidos; la fórmula para los pesos depende del modelo. Por ejemplo, en un metaanálisis de ensayos aleatorizados del efecto de un inhibidor de la enzima convertidora de la angiotensina (ECA) sobre la mortalidad en pacientes con arteriopatía coronaria, el resumen del efecto en el modelo de efectos fijos fue el riesgo relativo medio ponderado, ponderado por el inverso de la varianza del riesgo relativo de cada uno de los estudios incluidos. Véanse también *Modelo de efectos fijos* y *Modelo de efectos aleatorios*.

Revisión experta. Revisión de un protocolo, una propuesta o un manuscrito por personas con el mismo nivel de experiencia que el investigador que preparó estos documentos. Por ejemplo, las propuestas enviadas para su financiación por los NIH son sometidas a un proceso de revisión experta en la que científicos del mismo ámbito puntúan el protocolo utilizando criterios bien definidos. De manera similar, los manuscritos enviados a revistas médicas son sometidos a revisión experta por científicos que ayudan a los editores de la revista a decidir si se debe publicar el manuscrito.

Revisión sistemática. Revisión de la bibliografía médica en la que se utiliza un abordaje sistemático para identificar todos los estudios sobre una pregunta de la investigación determinada, unos criterios claros para incluir un estudio en la revisión, y métodos estandarizados para extraer los datos de los estudios incluidos. En una revisión sistemática también se puede incluir un metaanálisis de los resultados del estudio. Por ejemplo, el investigador realizó una revisión sistemática de todos los estudios en los que se evaluó si los suplementos de cinc reducen el riesgo de presentar catarro.

Riesgo relativo. Véase *Cociente de riesgos*.

Sensibilidad. Proporción de pacientes con la enfermedad en los que es positiva una prueba («positiva en la enfermedad» o PE). Por ejemplo, en comparación con los resultados anatomopatológicos de la biopsia, la sensibilidad de un resultado > 4,0 ng/ml en un análisis del PSA sérico es de aproximadamente el 20 % para la detección de cáncer de próstata; en otras palabras, el 20 % de los hombres con cáncer de próstata tendrán un PSA > 4,0 ng/ml. Véase también *Especificidad*.

Sesgo. Error sistemático en la medición, o en una asociación estimada, debido a un problema del diseño, la ejecución o el análisis del estudio. Por ejemplo, debido al sesgo en la forma en la que los participantes recordaron su exposición a productos químicos tóxicos, los pacientes con leucemia tenían más probabilidad de referir el uso de insecticidas que los testigos.

Sesgo de ausencia de respuesta. Tipo de sesgo en el que la ausencia de respuesta (p. ej., a un cuestionario) afecta a los resultados del estudio. Por ejemplo, los investigadores estaban preocupados por el sesgo de ausencia de respuesta en su estudio de los efectos del consumo de drogas sobre el riesgo de presentar insuficiencia renal.

Sesgo diferencial. Término general para una situación en la que la medición varía sistemáticamente según el estado del participante, habitualmente dependiendo de que el participante sea un caso o un testigo; se produce la mayoría de las veces con exposiciones que se recuerdan. Por ejemplo, como los casos de enfermedad celíaca del adulto tenían más probabilidades de evocar exposiciones a productos que contenían trigo durante la infancia que sus hermanos que habían crecido en el mismo hogar, los investigadores sospecharon que había sesgo de recuerdo diferencial. Véase también *Sesgo no diferencial.*

Sesgo no diferencial. Tipo de sesgo que no depende de que el participante sea un caso o un testigo (o en ocasiones de que un participante haya estado expuesto o no a una tercera variable). El sesgo no diferencial tiende a hacer que sea más difícil encontrar asociaciones porque reduce las diferencias aparentes entre los grupos. Por ejemplo, aunque el recuerdo de la exposición previa a antibióticos fue imperfecto en los casos y en los testigos, aparentemente el sesgo no fue diferencial, porque una revisión de las historias clínicas indicó que los dos grupos tenían inexactitudes similares. Véase también *Sesgo diferencial.*

Sesgo de doble método de referencia. Véase *Sesgo de verificación diferencial.*

Sesgo de espectro. Situación en la que la exactitud de una prueba es diferente en la muestra de lo que habría sido en la población porque el espectro de la enfermedad (que afecta a la sensibilidad) o de ausencia de enfermedad (que afecta a la especificidad) en la muestra difiere del de la población en la que se utiliza la prueba. Por ejemplo, debido al sesgo del espectro, se encontró que un nuevo análisis sérico diseñado para diagnosticar el cáncer esofágico era relativamente exacto en un estudio de pacientes con cáncer de esófago avanzado en comparación con estudiantes de medicina sanos, aunque tuvo un rendimiento bajo cuando se utilizó en pacientes ancianos con dificultad deglutoria no diagnosticada.

Sesgo de muestreo. Error sistemático que hace que la muestra de personas incluidas en un estudio no represente a la población objetivo. Por ejemplo, si se incluyera a los participantes en un estudio de factores de riesgo de osteoporosis de entre los pacientes hospitalizados por fractura de cadera, parecería falsamente que la caída es un factor de riesgo de osteoporosis debido al sesgo de muestreo.

Sesgo del observador. Situación en la que un investigador (o auxiliar de la investigación) realiza una evaluación no objetiva que depende de su conocimiento de uno o más atributos del participante, como si el participante es un caso o un testigo, o si ha estado expuesto o no a un factor de riesgo particular. Por ejemplo, aparentemente el sesgo del observador fue responsable del hallazgo de que, de acuerdo con la entrevista, había más probabilidad de considerar que tenían problemas de manejo de la ira los adolescentes hispanos que los asiáticos, porque en una encuesta autoadministrada y una revisión de los registros escolares no se encontraron diferencias entre ambos grupos.

Sesgo de participante. Véase *Sesgo de recuerdo*.

Sesgo de publicación. Distorsión de la bibliografía publicada que se produce cuando los estudios publicados no son representativos de todos los estudios que se han realizado, habitualmente porque los resultados positivos se envían y publican con más frecuencia que los resultados negativos. Por ejemplo, los autores del metaanálisis sospecharon sesgo de publicación porque encontraron que se habían publicado seis estudios positivos pequeños, y tan sólo un estudio negativo extenso.

Sesgo de recuerdo. Tipo específico de sesgo en el que el hecho de que un participante recuerde la exposición o factor de riesgo, y cómo lo recuerde, depende de otro factor, especialmente de si el participante es un caso o un testigo. Por ejemplo, se pensó que el sesgo de recuerdo era el motivo por el que los casos de esclerosis lateral amiotrófica tenían más probabilidades de recordar la exposición a insecticidas que los testigos.

Sesgo de verificación (también denominado sesgo de estudio diagnóstico o sesgo de derivación). Sesgo en la evaluación de la exactitud de una prueba que se produce cuando se verifica selectivamente la enfermedad en los pacientes mediante el estudio con un método de referencia basado, en parte, en los resultados de la propia prueba que se está estudiando. Por ejemplo, si en un estudio de la exactitud de la percusión torácica para diagnosticar neumonía se incluyera únicamente a pacientes en los que se hubiera realizado una radiografía de tórax, y aquellos en los que se detectara matidez a la percusión tuvieran mayor probabilidad de que se les realizara una radiografía, habría un falso aumento de la sensibilidad de la percusión, y una disminución falsa de su especificidad, debido al sesgo de verificación.

Sesgo de verificación diferencial. Sesgo que se produce en estudios de pruebas diagnósticas cuando se aplican métodos de referencia distintos a diferentes participantes, dependiendo, al menos en parte, del resultado de la prueba que se estudia. Por ejemplo, en un estudio

de cribado con el antígeno prostático específico (PSA) para detectar cáncer de próstata en hombres, a los que tenían concentraciones elevadas de PSA se les realizó una biopsia prostática, mientras que se realizó seguimiento clínico a los que tenían concentraciones normales de PSA; esto planteó la preocupación de que el sesgo de verificación diferencial hubiera aumentado falsamente la sensibilidad, y hubiera reducido la especificidad, del cribado mediante PSA en hombres con cáncer de próstata indolente.

Sujeto. Véase *Participante*.

Supresión. Tipo de confusión en la que el factor de confusión reduce la asociación aparente entre la variable predictiva y la variable de respuesta, porque se asocia a la variable predictiva, pero afecta a la variable de respuesta en la dirección *opuesta*. Por ejemplo, se podría pasar («suprimir») por alto una asociación entre el tabaquismo y las arrugas cutáneas si los fumadores fueran más jóvenes y no se controlara la confusión por la edad.

Tabla de datos. Tabla de los datos del estudio en la que cada hilera corresponde a un registro único y cada columna corresponde a un campo o atributo. En todos los estudios habrá una tabla de participantes en el estudio, cada una de cuyas filas corresponde a un participante individual, y las columnas corresponden a información específica del participante, como sexo y fecha de nacimiento. En la mayoría de los estudios se utilizarán tablas adicionales en las que las filas corresponden a visitas del estudio, resultados de laboratorio, contactos telefónicos, etcétera.

Tamaño de la muestra. Este término tiene dos significados. Puede ser el número de participantes incluidos en un estudio, o el número estimado de participantes que es necesario para que un estudio tenga éxito. Por ejemplo, el investigador estimó que necesitaría tener un tamaño de la muestra de 54 participantes para tener una potencia del 90 % a fin de detectar un aumento al doble del riesgo de conducta agresiva en niños de tercer curso expuestos a videojuegos violentos.

Tasa de incidencia. La tasa con la que se produce una enfermedad o una variable de respuesta particular en un grupo de participantes que previamente no tenían ese trastorno. Habitualmente se calcula como el número de nuevos casos de la variable de respuesta dividido por el número de persona-tiempo en riesgo. Por ejemplo, la tasa de incidencia de infarto de miocardio fue de 35,3 por cada 1 000 persona-años en hombres de mediana edad, y aproximadamente la mitad (17,4 por cada 1 000 persona-años) en mujeres de mediana edad. También se refiere a la medida del riesgo que se define como el número de personas que presentan una respuesta dividido por el número de personas-tiempo en riesgo. Por ejemplo, la tasa de incidencia de accidente cerebrovascular en el estudio fue de 23 por cada 1 000 personas-año. Véanse también *Cociente de riesgos instantáneos y Persona-tiempo*.

Tasa de respuesta. Proporción de participantes elegibles que responden a un cuestionario o a una pregunta particular del mismo. Una tasa de respuesta baja puede reducir la validez interna del estudio y sesgar los resultados. Por ejemplo, en una encuesta de estudiantes de secundaria, una tasa de respuesta del 20 % a una pregunta sobre el consumo de marihuana indicaría que es poco probable que el resultado sea una estimación válida de la frecuencia real de consumo de marihuana en estudiantes. Véase también *Datos ausentes*.

Tasa de riesgos instantáneos. Término epidemiológico que mide la incidencia instantánea a la que se produce una variable de respuesta en una población. En la práctica casi siempre se estima como la incidencia de una variable de respuesta. Por ejemplo, se estimó que la tasa de riesgos instantáneos de presentar arteriopatía coronaria en mujeres de 50 a 59 años de edad era de 0,008 al año.

Testigo. Término que tiene dos significados diferentes. Primero, testigo se refiere a un participante que no tiene la variable de respuesta de interés, por lo que es un miembro de un grupo de comparación con el que se comparan aquellos que sí tienen la variable de respuesta (los «casos»). Por ejemplo, en un estudio de factores de riesgo de enfermedad ulcerosa péptica, se seleccionaron los testigos de entre los pacientes hospitalizados durante el período del estudio con un diagnóstico no relacionado con el tubo digestivo. Segundo, testigo se refiere al «tratamiento» inactivo (p. ej., el placebo o el «tratamiento habitual») que recibieron los participantes de un ensayo clínico que no recibieron la intervención en estudio; en ese contexto también se utiliza testigo para referirse a un participante que recibió el tratamiento inactivo. Por ejemplo, a los testigos se les administraron comprimidos de placebo que tenían un aspecto idéntico al fármaco activo. Véanse también *Caso* e *Intervención*.

Testigo con placebo. Testigo inactivo que es indistinguible del fármaco activo o de la intervención utilizados en un ensayo aleatorio. Por ejemplo, en un ensayo aleatorio controlado con placebo de un nuevo tratamiento para la incontinencia, el placebo debe tener un aspecto, un olor, un sabor y un tacto iguales a los de la nueva medicación que se está estudiando.

Testigo procedente de la consulta. En el contexto de un estudio de casos y testigos, la selección de los testigos de las mismas consultas (o centros) de las que se extrajeron los casos. Por ejemplo, el investigador utilizó testigos procedentes de la consulta en su estudio de si correr sobre la acera al menos 3 km a la semana se asociaba a artrosis de rodilla en la radiografía.

Testigos procedentes del hospital. En un estudio de casos y testigos, la selección de los testigos de los mismos hospitales de los que se extrajeron los casos. Por ejemplo, en su estudio de si el consumo de carnes procesadas se asociaba a cáncer del tubo digestivo superior,

el investigador utilizó testigos procedentes del hospital seleccionados de entre los pacientes que tenían enfermedades no malignas del tubo digestivo tratados en el mismo hospital que los casos.

Validez. Grado en el que una medición representa el fenómeno de interés. Por ejemplo, la puntuación de un cuestionario de calidad de vida es válida en la medida en que realmente mide la calidad de vida.

Validez aparente. Término que describe en qué grado una medida parece medir un fenómeno particular, basado en si parece razonable; generalmente no es un método muy fiable para evaluar la validez. Por ejemplo, se consideró que una medición de la popularidad en adolescentes tenía validez aparente porque los investigadores pensaban que diferenciaba a los estudiantes populares de sus institutos de aquellos que no lo eran. Véanse también *Validez de constructo, Validez de contenido* y *Validez relacionada con el criterio*.

Validez de constructo. Término que describe en qué grado una medición corresponde a las definiciones teóricas del rasgo (el «constructo») que se está midiendo. Por ejemplo, se pensaba que una medida de ansiedad social tenía validez de constructo porque había diferencias importantes de sus valores en las personas cuyos amigos las describían como «amante de la diversión» y «extrovertido», y aquellas a las que describían como «tímido» y «poco probable que vaya a fiestas». Véanse también *Validez de contenido* y *Validez relacionada con el criterio*.

Validez de contenido. Término que describe en qué grado una medición representa varios aspectos del fenómeno que se está estudiando. Por ejemplo, se pensó que una medida del insomnio tenía validez de contenido porque medía la cantidad total de sueño, los episodios de despertares nocturnos, el despertar a primera hora de la mañana, la energía al levantarse por la mañana y la somnolencia diurna. Véanse también *Validez de constructo* y *Validez relacionada con el criterio*.

Validez predictiva. Término que describe en qué grado una medición representa el fenómeno subyacente que pretende medir, de acuerdo con su capacidad de predecir variables de respuesta relacionadas. Por ejemplo, la validez predictiva de la medición de la depresión se reforzaría si se asociara al riesgo de suicidio posterior.

Validez relacionada con el criterio. Término que describe en qué grado una medición se correlaciona con otras formas de medir el mismo fenómeno. Por ejemplo, se pensó que una medición de la depresión en adolescentes tenía validez relacionada con el criterio porque tenía una correlación elevada con las puntuaciones del cuestionario de depresión de Beck. Véanse también *Validez de constructo* y *Validez de contenido*.

Valor de***p.*** De acuerdo con las pruebas estadísticas, la probabilidad de encontrar un efecto (de manera más precisa, un valor del estadístico de prueba) tan grande o mayor que el que se encontraría en el estudio únicamente por el azar si la hipótesis nula realmente fuera cierta. Por ejemplo, si la hipótesis nula es que beber café no se asocia al riesgo de infarto de miocardio, y en el estudio se encontró que el riesgo de infarto de miocardio en bebedores de café en comparación con los no bebedores era 2,0 con un valor de *p* de 0,10, habría habido una probabilidad del 10 % de encontrar un riesgo relativo de 2,0 o mayor en este estudio si no hubiera una asociación entre el consumo de café y el infarto de miocardio en la población.

Valor predictivo negativo. Probabilidad de que una persona con un resultado negativo de una prueba no tenga la enfermedad que se estudia. Por ejemplo, en una población de hombres con una prevalencia de cáncer de próstata del 10 %, el valor predictivo de un antígeno prostático específico (PSA) ≤ 4,0 ng/ml es de aproximadamente el 91 %. Véanse también *Prevalencia, Probabilidad previa, Sensibilidad* y *Especificidad*.

Valor predictivo positivo. Probabilidad de que una persona con un resultado positivo de una prueba tenga la enfermedad que se estudia. Por ejemplo, en una población de hombres con una prevalencia de cáncer de próstata del 10 %, el valor predictivo positivo de un antígeno prostático específico (PSA) > 4,0 ng/ml es de aproximadamente el 30 %. Véanse también *Prevalencia, Probabilidad previa, Sensibilidad* y *Especificidad*.

Variabilidad. Magnitud de la dispersión de una medición, que habitualmente se calcula como la desviación típica. Por ejemplo, si el cambio del peso corporal que produce una dieta varía desde un aumento de peso importante hasta una pérdida de peso importante, el cambio es muy variable. Véanse también *Desviación típica* y *Error típico de la media*.

Variable. Medición que puede tener valores diferentes. Por ejemplo, el sexo es una variable porque puede adoptar dos valores diferentes, masculino y femenino. Véanse también *Variable categórica, Variable de confusión, Variable continua, Variable dicotómica, Variable discreta, Variable nominal, Variable ordinal, Variable de respuesta* y *Variable predictiva*.

Variable categórica. Variable que puede tener sólo varios valores posibles. Por ejemplo, el investigador transformó las mediciones del nivel educativo referidas en una variable categórica con cuatro valores: menos de secundaria, secundaria o algunos años de universidad, título universitario o título de posgrado. Véanse también *Variable continua, Variable dicotómica, Variable nominal* y *Variable ordinal*.

Variable de confusión. Véase *Confusión*.

Variable continua. Medición que, en teoría, puede tener un número infinito de posibles valores. En la práctica el término muchas veces se utiliza para mediciones que tienen «muchos» (según algunos autores 10 o más, y según otros 20 o más) valores posibles. Por ejemplo, la presión arterial sistólica se mide como variable continua en mm Hg utilizando un esfigmomanómetro de mercurio. Véanse también *Variable categórica, Variable dicotómica y Variable discreta*.

Variable dependiente. Véase *Variable de respuesta*.

Variable dicotómica. Variable que puede tener sólo uno de dos valores, como sí/no u hombre/mujer. Por ejemplo, el evaluador dicotomizó la presión arterial sistólica en hipertensa (≥ 140 mm Hg) o no. Véanse también *Variable categórica* y *Variable continúa*.

Variable discreta. Tipo de variable que adopta únicamente valores enteros. Con fines prácticos, las variables continuas en ocasiones se tratan como variables discretas. Por ejemplo, la edad generalmente se expresa como la edad en años en el último cumpleaños, y el tabaquismo actual como el número medio de cigarrillos fumados al día. Véase también *Variable continua*.

Variable independiente. Véase *Variable predictiva*.

Variable instrumental. Variable que se asocia con la variable predictiva, pero que no se asocia de ninguna otra manera con la variable de respuesta; por lo tanto, se puede utilizar para estimar indirectamente el efecto de la variable predictiva sobre la variable de respuesta. Por ejemplo, los investigadores encontraron grandes diferencias regionales en el uso de una nueva vacuna antigripal, por lo que pudieron utilizar la región de residencia como variable instrumental para estudiar el efecto de la vacuna antigripal sobre la mortalidad total en ancianos.

Variable nominal. Variable categórica para la cual no hay un orden lógico. Por ejemplo, la filiación religiosa (cristiano, budista, hindú, musulmán, judío, otra, ninguna) se codificó como variable nominal.

Variable ordinal. Variable categórica cuyos valores tienen un orden lógico. Por ejemplo, se trató el consumo actual de alcohol como una variable ordinal: los valores eran ausencia de consumo de alcohol, 1 o 2 bebidas a la semana, > 2 y < 7 bebidas la semana, 1 a 2 bebidas al día, y ≥ 3 bebidas al día. Véase también *Variable nominal*.

Variable predictiva. Cuando se considera la asociación entre dos variables, la que se produce primero o tiene más probabilidad, por motivos biológicos, de producir la otra. Por ejemplo, en un estudio para determinar si la obesidad se asocia a un aumento del riesgo de apnea del sueño, la obesidad sería la variable predictiva. En un ensayo aleatorizado analizado mediante intención de tratar, la variable predictiva es la asignación a un grupo determinado.

Variable de respuesta. Definición formal de la respuesta para cada uno de los participantes. Por ejemplo, en un estudio de los efectos de diferentes tipos de ejercicio sobre el peso corporal y la composición corporal, las variables de respuesta se definieron como el cambio del peso en kilogramos desde el momento inicial hasta la medición final después de 1 año, y el cambio de la circunferencia de la cintura en centímetros durante ese mismo período.

Variables categóricas policotómicas. Variables categóricas con tres o más categorías. Por ejemplo, el grupo sanguíneo, que incluye los grupos A, B y O, es una variable categórica policotómica.

Índice alfabético de materias

A
AACR.*V.* Análisis de árbol de clasificación y regresión (AACR)
AAG.*V.* Acontecimientos adversos graves (AAG)
Abandonos, 60 71
estudios de cohortes, 95
Abordaje(s)
alternativos, 284
bayesiano, 52
estadístico de los estudios de cohortes, 93 94 95
de tipo Bonferroni para hipótesis múltiples, 50 51
Aceptabilidad biológica, 121 131
Acontecimientos adversos
graves (AAG), 142
infrecuentes, 194

Acrónimo FINER, 3 17 17
Actitud escéptica, 15
Actividades clínicas, 262
Acuerdo interobservador, 188
Adjudicación de las respuestas, 162
Administrador, 281
científico, 278
ADN, 145 196
AIDSLINE, 198
Ajuste, 127 128
ejemplo simplificado, 135 136
excesivo, 173 181
multivariante, 61 62
Aleatorización, 145 146
adaptativa, 146
por bloques, 146
estratificados, 146
conglomerados, 152 153
grupos, 152
mendeliana, 125
pares emparejados, 147
técnicas, 146 147
α (alfa), 48 49 48
abordaje de Bonferroni, 168 169
pruebas múltiples, 164
Ámbito del estudio, 18
Análisis
de árbol de clasificación y regresión (AACR), 180 181
combinados, 202
«cómo se les trató», 165
de datos secundarios, 192 193 194 195 196 203
conjuntos de datos
agregados, 194
de base comunitaria, 194 195
individuales, 193
pregunta de la investigación, 195 196
ventajas y desventajas, 192
estadístico, 164 165 166
estudio(s)
efectos sobre los resultados, 185
pruebas
exactitud, 177 178 179
reproducibilidad, 173 174 175
resultados, efectos sobre las decisiones clínicas, 183
viabilidad, costes y riesgos, 182 183
κ para medir el acuerdo interobservador, cálculo, 188 188
revisión sistemática, 200 201
de intención de tratar, 161 164 165 185
intermedios, 154 155 168
de regresión logística, 62 180 180
según protocolo, 164 165
de subgrupos, 165 166 202 203
de supervivencia, 61 179
Anatomía de la investigación, 2 3 4 5 6
Apéndices de una propuesta, 284
Aplicación del estudio, 7 8 8
Apoyo
corporativo, 288 289 290

interno, 290
universitario, 290
Archivo informático, 16
Asignación
aleatoria, 145
de formato, cuestionario, 225 226 227
Asistencia a conferencias, 15
Asociación(es)
espuria, 117 118 119 120 121 118 120
medidas para el cálculo, 111 112 113
reales distintas a la relación causa-efecto, 121 122 121
solidez, 131
Aspectos
estadísticos
ajuste(s)
factores de confusión, 128
multivariante, 61 62
datos ausentes, 261
emparejamiento, 124
hipótesis, 46 47 48 49 50
investigación clínica, 6
políticos, 268
Aumento de la precisión, estrategias, 37 41
Ausencia
de inferioridad, 154 167
ensayos, 62 63 139 153 154 167
margen (Δ)
de tratamiento activo, 139
Autoría
aspectos éticos, 218
fantasma, 218
honorífica, 218
invitada, 218
requisitos, 289
Azar, asociaciones espurias, 117 118 119 118
B
Barreras
a la comunicación en estudios internacionales, 271 272
culturales en estudios internacionales, 271 272
por la distancia en estudios internacionales, 271 272
lingüísticas, estudios internacionales, 271 272
Base(s) de datos
administrativa electrónica, 193
clínicos, 193
computarizada, 193
electrónicas, 247
del estudio, 248
relacional, 239 244 246
con múltiples tablas, 238
Beca(s)
contratos, 286
financiación privada, 288 289
fundación(es), 288 289
y sociedades especializadas, 288 289
National Institutes of Health, 286 287 288 287 288
Beneficencia, principio, 210
Beneficio, interrupción, 163
β (beta), 48 49 48

Bibliografía de la propuesta, 284 285
Bonferroni, método, 168 169
Brecha 10/90, investigación sanitaria, 268
Buena(s)
comunicación y compromiso a largo plazo, 274
pregunta de investigación, 17 18 19
propuestas, características, 285
Búsqueda, 246
extensa, 70
C
Calendario
establecimiento, redacción de propuesta, 278
realización del proyecto de investigación, 284 284
Calibración
formación, certificación, 258
instrumento, 37
Calidad
científica, 285
de vida, 34
Cambio
menor, 212 255
de mentor, 16
Campo, 237 238
CANCERLIT, 198
Capacidad
administración, 273
investigación, 268 274
Captura electrónica de datos, introducción de datos, 242 243
Cardiopatía isquémica (CI), 4 25 57 58 139 140
Catástrofes, 275
Causalidad, 131 132
Causas-efecto, 117 118 121 132
Censo o registro, 92
Center for Scientific Review (CSR), 286
Centro coordinador, 262
Certificación
calibración, formación, 258
observador, 36
Certificado de defunción, registro, 193
CI.*V.* Cardiopatía isquémica (CI)
Cierre del estudio, 256 257
Claridad
enunciado de un cuestionario, 227
presentación, 285
Clave principal, 237
Cociente(s)
de riesgos (riesgo relativo), 59 93 94 93
instantáneos
pruebas pronósticas, 179
de tasas, 93 94
de verosimilitudes, 177 178 179 179
Códigos de barras, 259
Coeficiente
de correlación (*r*), 35 59 60 79
de variación (CV), 34 175
Cointervención, 139 147
Colaboración
estudios internacionales, 271 272 274

investigación comunitaria, 270 271
modelo
ascendente, 271
descendente, 271
transcultural, 270 271 274
Comité(s)
directivo, 192 262
de revisión
estudio internacional, 273
ética en ambos países, 273
institucional (CRI), 253
aspectos éticos de la pregunta de la investigación, 18 19
autorización, 211 212 254
excepciones, 211 212 211 212
de supervisión de los datos y la seguridad (CSDS), 163 219 220
Compañero, 195 251 270
Compensación, 11
del participante en una investigación, 221
Completos y mutuamente excluyentes, 224 243
Comprensión por los participantes, 213
Comunicación de conflictos de intereses, 219
Comunidad
conjuntos de datos, 194 195
estudios, 268 269 270 271 272 273 274 275 276 269
investigación, 268 270 271
participación, 270
población, 27
Concordancia, 125
Condicionamiento, 130
éxito compartido, 129 130
Conducta científica impropia, 217 218
Confidencialidad, 211 215
base de datos, 247 248
certificado, 215
violación, 215
Conflictos de interés, 219 220
económicos, 219
Confusión, 117 118 121 130
aleatorización al tratamiento, 184
conjuntos de datos combinados, 194
fase(s)
análisis, 126 127 128 129 127
diseño, 122 123 124 125 126 123
indicación del tratamiento, 128 184
variable, 61 62 122
abordaje, 122 123 124 125 126 127 128 129 123 127
estudio de cohortes múltiples, 93
Congruencia
causalidad, 131
control de calidad, 257 258 259 260 261 262 263
calibración, formación y certificación, 258
coordinador de control de calidad, 257 258
datos
ausentes, 261
inexactos e imprecisos, 261 262
estudios multicéntricos colaborativos, 262
gestión de datos, 260 261 262 260 261
informes periódicos, 259

manual operativo, 258
práctica clínica correcta, 158 257 257
procedimientos
especiales para intervenciones medicamentosas, 259
de laboratorio, 259 260
revisión del rendimiento, 258 259
escala para medir variables abstractas, 230
precisión, 34
Conjunto(s) de datos
accesibles por métodos electrónicos, 27
agregados, 194 195
existentes
análisis de datos secundarios, 192 193 194 195 196
conjuntos de datos
de base comunitaria, 194 195
combinados, 194
individuales, 193
pregunta de la investigación, 195 196
ventajas y desventajas, 192
estudios auxiliares, 196 197
revisiones sistemáticas, 197 198 199 200 201 202 203 198
análisis de subgrupos y de sensibilidad, 202 203
criterios para la inclusión y exclusión de estudios, 199 199
evaluación del sesgo de publicación, 201 202 202
metaanálisis, 200 201
pregunta de la investigación, 198
recogida de datos, 199 200 200
individuales, 193
nacionales, 193
de validación, 181
Consentimiento
compresión por los participantes de la información desvelada, 213
formularios, 213
informado, 211 212 213 214 215 215
naturaleza voluntaria, 213
voluntario, 212 213 214 215
Consistencia interna, 230
Consulta
desarrollo del plan del estudio, 19
propuesta, 284
Consultor, 17 273 285
Consumibles, 253 254
Contratos, 253 289
laboratorios comerciales, 260
National Institutes of Health, 286 287 288 287 288
Control
análisis y publicaciones, 219
de calidad, 257 258 259 260 261 262 263
calibración, formación y certificación, 258
coordinador, 257 258
datos
ausentes, 261
fraudulentos, 262
inexactos e imprecisos, 261 262
estudios multicéntricos colaborativos, 262
gestión de datos, 260 261 262 260 261
informes periódicos, 259
manual operativo, 258

procedimientos
especiales para intervenciones farmacológicas, 259
de laboratorio, 259 260
revisión del rendimiento, 258 259
Copia de seguridad de datos, 248
Corporaciones, 286 289
Costes
ensayo aleatorizado con enmascaramiento, 143
estudio, 182 183
tiempo y dinero, 18
Cox, análisis
regresión, 94
riesgos proporcionales, 62 94 164 180
CRD.*V.* Cuaderno de recogida de datos (CRD)
CRDe.*V.* Cuaderno de recogida de datos electrónico (CRDe)
Creatividad, origen de la pregunta de la investigación, 16
CRI.*V.* Comité de revisión institucional (CRI)
Criterio(s)
de exclusión, 5 26 27
ensayo aleatorizado con enmascaramiento, 143 144 143
estudios de pruebas diagnósticas, 185 186
revisión sistemática, 198
de inclusión, 26 27 26
ensayo aleatorizado con enmascaramiento, 142 143 144
revisión sistemática, 198
de selección, 12 142 143 144 143
para participantes en estudios, 25 26 27 26
de valoración
compuestos, medición ensayos aleatorizados con enmascaramiento, 140 141
intermedios, 149
principal único, 140
Cronbach, a, 230
CSDS.*V.* Comité de supervisión de los datos y la seguridad (CSDS)
CSR.*V.* Center for Scientific Review (CSR)
Cuaderno de recogida de datos (CRD), 242
electrónico (CRDe), 242
Cuestionario(s), 223 224 225 226 227 228 229 230 231 232 233 234
asignación de formato, 225 226 227
autoadministrado, 234
comparación con entrevista, 232
por correo electrónico, 233
determinación del marco temporal, 227 228
discordancia entre la pregunta y las opciones de respuesta, 229
enunciado, 227
escalas y puntuaciones para medir variables abstractas, 229 230
métodos de administración, 233
páginas web, 233
pasos en la recogida de instrumentos para el estudio, 230 231 232
preguntas
abiertas y cerradas, 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243
de doble efecto, 228
suposiciones ocultas, 228 229
Cumplimiento del protocolo, 160 161 162 160
Curvas de eficacia diagnóstica (ROC), 177 177
CV.*V.* Coeficiente de variación (CV)
D
Datos
administrativos, 237

análisis, 247
ausentes, 261
base de datos, 237
almacenamiento fuera del centro, 248
análisis de datos secundarios, 193 195
auditoría, 248
búsqueda, 237
resultados, ictericia del lactante, 246
computarizada, 193
confidencialidad, 247 248
consola integrada, 245
copias de seguridad, 248
dos tablas, 238
ictericia del lactante, 239
electrónica, 247
identificadores personales, 238 248
integridad referencial, 239
normalización, 239
relación, 238
seguridad, 247 248
sistema
de apoyo, 243
frontal, 243
uso de hojas de cálculo, 237
búsquedas, 246 247
conjunto
combinados, 194
de base comunitaria, 194 195
individuales, 193
diccionario, 240
edición, 247
elementos, 241
errores, identificación y corrección, 246 247
extracción, 246 247
formularios electrónicos en pantalla, ventajas, 242
fraudulentos, 262
gestión
control de calidad, 260 261 262 260 261
estudio preliminar, 250
investigación clínica, 237 238 239 240 241 242 243 244 245 246 247 248
programas informáticos, 245
imprecisos o inexactos, 261 262
introducción, 242 243 244 245 246
captura electrónica, 242 243
distribuida, 242
edición, 248 262
formularios legibles por ordenador, 242
grupo de opción, 243 244
importación de mediciones, 243
lista
desplegable, 243 244
para escoger, 243 244
opciones de respuesta, 243
completas, 243
mutuamente excluyentes, 243
paquetes informáticos, 245
programas informáticos para la gestión de los datos, 243 244 245 246
respuesta(s)

opciones, 243
completas, 243
mutuamente excluyentes, 243
codificadas y texto libre, 243
resultados de laboratorio, 243
tabla, 240
sistema, 237
transcripción con el teclado, 242
preliminares, 283
tabla, 237 238 239 240 241 241
clave principal, 237
estudio de cohortes, 237 238
fichero plano, 237 238
número de identificación del participante, 237
tabla única, 237
tabulados, 259
tipos, 240
Declaración resumen, 287
Definición operativa, 255
Delatores, 218
Densidad mineral ósea (DMO), 140 193
Derivación, 181
Desviación típica, 34 70 73 175
Deterioro
de la capacidad de toma de decisiones, 209 210
cognitivo/comunicativo, 216
Diario, 228
Diazinon-oxón, 125
Diferencias
culturales, 268 272
de riesgos, 179
Diplomacia sanitaria, 275
mundial, 275
Diseño(s)
adaptativos, 154 155
casos y testigos con densidad de incidencia, 97
cruzamiento, 156 157 158 157
estudio(s), 6 7 6
abordaje, 3 4 5 4
análisis de datos secundarios, 192 193 194 195 196
conjuntos de datos
de base comunitaria, 194 195
combinados, 194
individuales, 193
pregunta de la investigación, 195 196
ventajas y desventajas, 192
auxiliares, 196 197
de casos y testigos, 97 98 99 100 101 102 103 104
eficiencia para variables de respuesta infrecuentes, 99
estructura, 97 98 99 98
generación de hipótesis, 100
medición diferencial, sesgo, 102 103 104 103
sesgo de muestreo, 100 101 102 100
de cohortes, 85 86 87 88 89 90 91 92 93 94 95
abordaje estadístico, 93 94 95
casos y testigos anidado con densidad de incidencia, 104 105 106 107 108 106
cohortes múltiples y testigos externos, 91 92 93 92
diseño de casos y testigos anidado, 104 105 106 107 108 105

estudio transversal, 85 86 87 88 86
prospectivo, 88 89 90 88
retrospectivo, 90 91 90
cruzados, 108
cumplimiento del protocolo, facilidad, 160 161 162 160
diseños observacionales, elección, 108 109
elaboración del protocolo, 10
ensayo aleatorizado con enmascaramiento, 147 148 149
aplicación de las intervenciones, 147 148 149 148
asignación aleatoria de los participantes, 145 146 147 148 149
intervención, elección, 137 138 139
medición(es) de las variables
iniciales, 144 145
de respuesta, 140 141 142
selección de los participantes, 142 143 144 143
testigo, elección, 139
ensayo clínico, 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169
aleatorización por conglomerados, 152 153
análisis de los resultados, 164 165 166
autorización
nuevas intervenciones por las autoridades de registro, 158 159
nuevos tratamientos por la FDA, 158 158
determinación y adjudicación de las variables de respuesta, 162
diseño(s)
adaptativos, 154 155
con cruzamiento, 156 157 158
factorial, 151 152
entre grupos no aleatorizados, 155
intragrupo, 155 156 156 157
ensayos con testigo activo, 153 154
estudios piloto, 159
método de Bonferroni, 168 169
seguimiento y cumplimiento del protocolo, 160 161 162 160
supervisión, 163 164 164
inferencia causal, 8 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136
asociaciones reales distintas a la relación causa-efecto, 121 122 121
errores en la cuantificación de los efectos causales, 129 130
espuria en diseños observacionales, 117 118 119 120 121 120
estrategia, elección, 130 131 132
fase(s)
análisis, factores de confusión, 126 127 128 129 127
diseño, factores de confusión, 122 123 124 125 126 123
minimización del sesgo, 119 120 121 120
pruebas diagnósticas y pronósticas, 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190
determinación de la utilidad del estudio, 171 172 173 172
errores de diseño, 185 186 187
exactitud de las pruebas, estudios, 175 176 177 178 179 177 192
κ para medir el acuerdo entre observadores, cálculo, 188 188
problemas, 171 172 173
pruebas, efectos sobre las variables de respuesta, 183 184 185
reproducibilidad de pruebas, estudios, 173 174 175
resultados de pruebas, efectos sobre las decisiones clínicas, 181 182
sesgo de verificación, 186 187
viabilidad, costes y riesgos de las pruebas, estudios, 182 183
reducción del sesgo, 219
reunión de participantes, 25 25
factorial, 151 152 152
entre grupos, 155 156 156 157

no aleatorizado, 155
del período de preinclusión, 162
de la preinclusión con placebo, 162
visual, 226
Dispositivos electrónicos, 228
portátiles, 233
Distribución de las respuestas, 40
DMO.*V.* Densidad mineral ósea (DMO)
Docencia, 16
Dominios, 240
Duplicados con enmascaramiento y medidas de consenso, 260
E
EAV.*V.* Escala analógica visual (EAV)
Edición computarizada, 261
Efectividad, 193 194
Efecto(s)
adversos, 141 142
de aprendizaje, 156
causales, infraestimación, 130
compartido, condicionamiento, 129 130
magnitud, 47 48
estandarización, 56 57
tamaño de la muestra fijo, 65
modificación, 26 122 123 128 133 134 143 151 165
resumen, 200 205
Efecto-causa, 117 121 121
Eficacia, 138 194
Eficiencia, 40
estudio de casos y testigos, 99
Ejecución del estudio, 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267
base de datos, diseño, 254 255
control de calidad, 257 258 259 260 261 262 263
calibración, formación y certificación, 258
datos
ausentes, 261
fraudulentos, 262
inexactos e imprecisos, 261 262
estudios multicéntricos colaborativos, 262
gestión de los datos, 260 261 262 260 261
informes periódicos, 259
manual operativo, 258
práctica clínica adecuada, 257 257
procedimientos
especializados para intervenciones farmacológicas, 259
de laboratorio, 259 260
estudio preliminar, 255
finalización, 256
reunión de los recursos, 251 252 253 254 255
equipo de investigación, 251 252
espacio, 251
liderazgo y formación del equipo, 252
puesta en marcha del estudio, 253 254
revisiones de protocolo, 255 256
Ejemplo numérico, sesgo de verificación, 189 190
Elección y dosis de la intervención, 149
Embarazada como participante en investigación, 217
EMBASE, 198
Embrión participante en la investigación, 217

Emparejamiento, 61 130
estructura, 2 3 4 5 6 3
aspectos estadísticos, 6
diseño del estudio, 3 4 5
introducción e importancia, 3
participantes en estudios, 5
pregunta de la investigación, 2 3 14 15 16 17 18 19 20 21
variables, 5
estudio de casos y testigos, 102
excesivo, 125
por frecuencia, 123
por parejas, 123
participantes en estudios, 5 23 24 25 26 27 28 29 30 31
criterios
de inclusión, 26 27 26
de selección, establecimiento, 26 27
diseño del protocolo para la adquisición, 25 25
generalización de los hallazgos del estudio, 24 25 24
muestras
de conveniencia, 27 28
probabilísticas, 28
número, 17
población(es)
clínicas y comunitarias, 27
y muestra objetivos, 24
resumen de las opciones de diseño del muestreo, 28 29
selección, 29 30
utilización de datos existentes, 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 220 221
análisis de datos secundarios, 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206
estudios auxiliares, 196 197
revisión sistemática, 198 198 202 202
ventajas y desventajas, 192
variables de confusión, 123 124 125 123
Encuesta(s), 223 224 225 226 227 228 229 230 231 232 233 234 235 236
comité de revisión institucional, 211 211 212
en línea, 223 226 227
seriada, 87 88
transversales seriadas, 88
Enmascaramiento
control de calidad, 259 260
estudios de pruebas médicas, 172 173
exactitud, 37
importancia, 147 148 149 148
imposibilidad, 148 149
sesgo, 147 148
de medición diferencial, 37 102 103 104 103
Enriquecimiento de la carrera profesional, 275
Ensayo(s)
aleatorizado con enmascaramiento, 3 4 137 138 139 140 141 142 143 144 145 146 147 148 149
alternativas, 151 152 153 154 155 152 156
aplicación de las intervenciones, 147 148 149 148
asignación aleatoria de los participantes, 145 146 147 148 149
criterios de valoración
clínicos, 140
compuestos, 140 141
intervención, elección, 137 138 139
medición(es), variables
iniciales, 144 145

de respuesta, 140 141 142
período de preinclusión previo, 162
pruebas diagnósticas, 184
selección de los participantes, 142 143 144 143
testigo, elección, 139
clínico(s), 3 4 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169
aleatorizado
aspectos éticos, 220
con enmascaramiento, 137 138 139 140 141 142 143 144 145 146 147 148 149
aplicación de las intervenciones, 147 148 149 148
asignación aleatoria de los participantes, 145 146 147 148 149
intervención, elección, 137 138 139
selección de los participantes, 142 143 144 143
testigo, elección, 139
variables iniciales, medición, 144 145
análisis de los resultados, 164 165 166
autorización por la FDA de nuevos tratamientos, 158 158
control de calidad, 257 258 259 258
criterios de exclusión, 27
diseños
aleatorizados alternativos, 151 152 153 154 155
adaptativos, 154 155
aleatorización por conglomerados, 152 153
ensayos con testigo activo, 153 154
factorial, 151 152
no aleatorizados, 155 156 157 158 159
autorización de nuevas intervenciones por autoridades de registro, 158 159
con cruzamiento, 156 157 158
estudios piloto, 159
entre grupos, 155
intragrupo, 155 156 156 157
evaluación y aplicación de las variables de respuesta, 162
método de Bonferroni, 168 169
problemas éticos, 209 210 211 212 213 214 215 216 217 218 219 220 221
autoría, 218
autorización por el comité de revisión institucional, 211 212 211 212
conducta ética impropia, 217 218
conflictos de interés, 219 220
consentimiento informado, 212 213 214 215 215
ensayos clínicos aleatorizados, 220
investigación en muestras y datos recogidos previamente, 220 221
principios éticos, 209 210
seguimiento y cumplimiento del protocolo, 160 161 162 160
supervisión, 163 164 164
variables de respuesta, efecto, 184 185
del equipo de investigación, 255
de equivalencia, 62 63 139 153 154
fase(s)
I, 158 158
II, 158 158
III, 158 158
IV, 158 158
para la Food and Drug Administration (FDA), 158 158
multicéntricos, 146 289
de «N de uno», 156
preclínico, 158 158
con testigo activo, 153 154
Entrevista(s), 223 224 225 226 227 228 229 230 231 232 233 234

asignación de formato, 225 226 227
comparación con cuestionario, 232
desarrollo, 230
determinación del marco temporal, 227 228
discordancia entre la pregunta y las opciones de respuesta, 229
enunciado, 227
escalas y puntuaciones para medir variables abstractas, 229 230
pasos para recoger instrumentos para el estudio, 230 231 232
en persona, 233
preguntas
abiertas y cerradas, 223 224 225
de doble efecto, 228
suposiciones ocultas, 228 229
telefónica, 233
asistida por ordenador (ETAO), 233
Enunciado del cuestionario, 227
Epidemiología
genética molecular, 41
molecular, genética, 41
Equilibrio, 155 220
Equipo
de investigación, miembros, funciones, 252
redacción de la propuesta, 277 278
revisión del rendimiento, 258 259
Error(es)
aleatorio, 9 9 118
precisión, 34 38
de clasificación, 103
diferencial, 103
no diferencial, 102 103
conceptual terapéutico, 212
de diseño, 7
frecuentes, 71 72
de la investigación, 8 9 10 9 10
sistemático, 9 9
exactitud, 37 38 41
minimización, 118 119
típico de la media (ETM), 73
de tipo I, 46 47 48 63
abordaje de Bonferroni, 168 169
minimización, 117 118 119
pruebas múltiples, 163 164
de tipo II, 46 47 48
Erudición, pregunta de la investigación, 15
Escala(s), 32 33 34 33
acortamiento, 231
analógica visual (EAV), 224
creación, 230
medición de variables abstractas, 229 230
con múltiples preguntas, 229
Escepticismo, 15 285
Espacio, 251
Especificación, variables de confusión, 122 123 123
Especificidad, 177 189
hipótesis, 44
medición, 39
tamaño de la muestra, 64
Estadística descriptiva, 28 29 63 183

Estrategia de investigación de la propuesta, 282 283 284 283
Estratificación, 126 127 143
Estudio(s), 182
analítico, 5
técnicas para el tamaño muestral, 55 56 57 58 59 60 56
coeficiente de correlación, 59 60 79
prueba(s)
de la ji al cuadrado (χ^2), 57 58 59 75
de la *t*, 56 57 73
de antes/después para la toma de decisiones clínicas, 182
auxiliares, 196 197 203
de casos
y cohortes, 104 105 106 107 108 105
y testigos, 3 97 98 99 100 101 102 103 104
anidado, 97 104 105 106 107 108 105 109
con densidad de incidencia, 104 105 106 107 108 106
cálculo de medidas de asociación, 111
efecto-causa, 121
eficiencia para respuestas infrecuentes, 99
emparejamiento, 123 124 125
estructura, 97 98 99 98
generación de hipótesis, 100
puntos
débiles, 100 101 102 103 104
fuertes, 99 100
razón de posibilidades como estimación del riesgo relativo, 114 115
sesgo
de medición diferencial, 102 103 104 103
de muestreo, 100 101 102 100
testigos
múltiples, uso, 69
procedentes de la consulta, 101
ventajas y desventajas, 109
de cohortes, 3 88 89 90 91 92 93 94 95
abordaje estadístico, 93 94 95
causa-efecto, 121
comparación con estudio de casos y testigos, 97
diseño de casos y testigos anidado, 104 105 106 107 108 105
con «densidad de incidencia», 104 105 106 107 108
estudio auxiliar, 196
incidencia, 89
múltiples, 91 92 93 92 109
y testigos externos, 91 92 93 92
problemas, 94 95 95
prospectivo, 88 89 90 88 89 109
prueba diagnóstica, 172 175
retrospectivo, 90 91 90 109
ventajas y desventajas, 109
de confirmación, 18
criterios de valoración, efecto, 183 184 185
cruzados, análisis, 108
descriptivo, 4
técnicas para el cálculo del tamaño muestral, 63 64 65 80 81
de doble cohorte, 91 92 92
ecológicos, 195
de eficacia comparativo, 153
inéditos, 201

internacionales, 268 269 270 271 272 273 274 275 276
aspectos éticos, 273 274 275 274
barreras de distancia, idioma y cultura, 271 272
colaboración, 271 272 274
justificación, 268 269 270 269
problemas de financiación, 272 273
recompensa, 275
riesgos y frustraciones, 275
nuevas terapias, fases, 158 158
observacionales, 3 4
comparación con ensayos clínicos., 137
efecto de la prueba sobre una variable de respuesta, 184
elección, 108 120
estudio de casos y testigos, 97 98 99 100 101 102 103 104
eficiencia para variables de respuesta infrecuentes, 99
estructura, 97 98 99 98
generación de hipótesis, 100
sesgo
de medición diferencial, 102 103 104 103
de muestreo, 100 101 102 100
estudio(s) de cohortes, 88 89 90 91 92 93 94 95
de casos y testigos anidado con densidad de incidencia, 104 105 106 107 108 106
diseño de casos y testigos anidado, 104 105 106 107 108 105
múltiples y con testigos externos, 91 92 93 92
prospectivo, 88 89 90 88
retrospectivo, 90 91 90
estudio transversal, 85 86 87 88 86 89
inferencia causal, 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136
asociaciones
espurias, 118 119 120 121 120
reales distintas a la relación causa-efecto, 121 122 121
errores en la cuantificación de los efectos causales, 129 130
estrategia, elección, 129 130 131 132
fase(s)
análisis, factores de confusión, 126 127 128 129 127
diseño, factores de confusión, 122 123 124 125 126 123
prueba(s)
de cribado, 184
diagnósticas, 171
oportunistas, 125 126
piloto, 17 70 159
estudio preliminar de los métodos del estudio, 255 256 257
preliminar, 232 255
previos, 14
de pruebas médicas, 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190
cálculo de κ para medir el acuerdo intraobservador, 188 188
determinación de la utilidad del estudio, 171 172 173 172
errores de diseño frecuentes, 185 186 187
estudio, efectos sobre los resultados, 183 184 185
exactitud de las pruebas, estudios, 175 176 177 178 179
cocientes de verosimilitudes, 177 178 179 179
curvas de rendimiento diagnóstico, 177 177
mejora neta de la clasificación, 179
muestreo, 175 176
riesgos absolutos, cocientes de riesgos, diferencias de riesgos y cocientes de riesgos instantáneos, 179
sensibilidad y especificidad, 177
variables
predictivas, 176

de respuesta, 176
problemas frecuentes, 171 172 173
reglas de predicción clínica, 180 181
reproducibilidad de las pruebas, estudios, 173 174 175
resultados de la prueba, efecto sobre las decisiones clínicas, 181 182
viabilidad, costes y riesgos de las pruebas, estudios, 182 183
de rendimiento diagnóstico, 182
de reproducibilidad, 173 174 175
de pruebas, 173 174 175
de series cronológicas, 155 156
tándem, prueba, 176
transversal(es), 3 85 86 87 88 86 89
cálculo de medidas de asociación, 111
causa-efecto, 121
coeficiente de correlación, 59 60
pruebas diagnósticas, 172
puntos fuertes y débiles, 86 87
ventajas y desventajas, 109
ETAO.*V.* Entrevista telefónica asistida por ordenador (ETAO)
Etiquetado, control de calidad, 259
ETM.*V.* Error típico de la media (ETM)
Evaluación de la heterogeneidad, 201
Exactitud, 32
enmascaramiento, 37
estrategias para mejorar, 37 38
investigación, 36 37 38 36 38
prueba, estudio, 175 176 177 178 179
cocientes de verosimilitud, 177 178 179 179
curvas de rendimiento diagnóstico, 177 177
mejora neta por la reclasificación, 179
muestreo, 175 176
riesgos absolutos, cocientes de riesgos, diferencias de riesgos y cocientes de riesgos instantáneos, 179
variables
predictivas, 176
de respuesta, 176
Exención de revisión por el CRI, 211
Experiencia
orígenes de una pregunta de investigación, 14 15 16
técnica, 17 18
Experimento(s)
ensayo aleatorizado con enmascaramiento, 137 138 139 140 141 142 143 144 145 146 147 148 149
aplicación de las intervenciones, 147 148 149 148
asignación aleatoria de los participantes, 145 146 147 148 149
intervención, elección, 137 138 139
selección de los participantes, 142 143 144 143
testigos, elección, 139
variables iniciales, medición, 144 145
naturales, 125
problemas éticos, 209 210 211 212 213 214 215 216 217 218 219 220 221
autoría, 218
autorización por el comité de revisión institucional, 211 212 211 212
conducta científica impropia, 217 218
conflictos de interés, 219 220
consentimiento informado, 212 213 214 215 215
ensayos clínicos aleatorizados, 220
investigación con muestras y datos recogidos previamente, 220 221
pago a los participantes en la investigación, 221
principios éticos, 209 210

técnicas para determinar el tamaño de la muestra, 55 56 57 58 59 60 56
coeficiente de correlación, 59 60 79
prueba(s)
de la ji al cuadrado (χ^2), 57 58 59 75
de la *t*, 56 57 73
Exposición, estudios de cohortes múltiples, 91 92 93
«Extensiones sin coste», 253
Extracción de datos, 246 247
F
Fabricantes de fármacos y dispositivos, 291
Factor(es)
de confusión, 126 127 128 129 127
constitucionales, emparejamiento, 124
Falacia ecológica, 195
Falsificación, 218
Fármaco activo, 63 138
Fase(s)
análisis, 120
diseño, 119
manejo de los factores de confusión, 122 123 124 125 126 123
en el estudio de nuevos tratamientos, 158 158
Fecha de inclusión, 104 106
Fenómenos de interés, 7 10
diseño de mediciones, 32 32 33
fisiología, de la investigación, 2
Fichero plano, 237 238
Filas, 237
Filtro, 246
Financiación
agencia
directrices, 278
envío de propuesta, 277
estudios internacionales, 272 273
limitaciones, 268
propuesta, 285 286 287 288 289 290 291
apoyo
corporativo, 289 290
interno, 290
becas
y contratos del NIH, 286 287 288 287 288
de fundaciones y sociedades especializadas, 288 289
Flexibilidad, 126
Fondos internos, 290
Forma de campana/embudo, 201
Formación
calibración y certificación, 258
centralizada, 262
observador, 36
Fórmula del tamaño de la muestra total, 78
Formulario(s)
consentimiento informado, 213
en papel, 226 242
Fuentes de variabilidad, 41
Fuerza de la asociación, 131
Funciones dobles de médicos-investigadores, 219
Fundaciones y sociedades, 290
Futilidad, interrupción, 163

G
Gestión económica, 272
Gestor(es)
de datos clínicos, 237
antes de un galardón, 253
después de un galardón, 253
del proyecto, 278
Gráfico(s)
de embudo, 201 202
organizativo, 284
Grupo(s)
de comparación, 68 273
adecuado, 273
dirigido, 230
testigo múltiples, 102
H
Hallazgos analíticos, 200
Health Insurance Portability and Accountability Act (HIPAA), 215 216 247
Heterogeneidad, 200 201 202
HIPAA.*V.* Health Insurance Portability and Accountability Act (HIPAA)
Hipótesis, 6 43 44 45 46
a posteriori, 45 50 51 52
alternativa(s), 45
bilateral, 45 46
pruebas estadísticas unilaterales y bilaterales, 45 49 50 55
unilateral, 45 46 58
características, 44 45
compleja, 44
específica, 44
generación, 52 100
múltiples, 50 51 52
y *a posteriori,* 50 51 52
problema del estudio, 163
nula, 45
estudio de equivalencia, 62 63
hipótesis alternativa, 45
α, β y potencia, 48 49 48
supervisión intermedia, 168
valor de *p,* 49
preliminar, 44 45
principal, 52 53
secundarias, 52 53
sencilla, 44
Historias clínicas electrónicas, 183
Hoja de cálculo, 248
Homogeneidad, 201 206
I
Imaginación, origen de la pregunta de la investigación, 16
IMC.*V.* Índice de masa corporal (IMC)
Importancia de la pregunta de la investigación, 19
Incertidumbre, pregunta de la investigación, 14
Incidencia
acumulada.*V.* Incidencia
estudio transversal, 86 88
Indefinición, 44
Índice
de masa corporal (IMC), 246
de mortalidad nacional, 193

Inferencia(s), 6
causal, 8 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136
asociaciones
espurias, 117 118 119 120 121 118 120
reales distintas a la relación causa-efecto, 121 122 121
errores en la cuantificación de los efectos causales, 129 130
estrategia, elección, 129 130 131 132
fase(s)
análisis, factores de confusión, 126 127 128 129 127
diseño, factores de confusión, 122 123 124 125 126 123
errónea, 9
estudios de cohortes, 94
generalización de los participantes en estudios a las poblaciones objetivo, 24
Influencia indebida, 213
Información
insuficiente, tamaño de la muestra, 55
privada, 210
sanitaria protegida, 144 215
Informe(s), 281
del investigador, 285
periódicos, 259
Instrumento(s)
automatización, 36
exactitud, 37
existentes, 231
mejora, 36
sesgo, 37
variabilidad, 34
Integridad
ética, 221
referencial, 239
Interacción.*V.* Modificación del efecto
Intervalo de confianza, 63 200
revisión sistemática, 200 205
del 95%, 130 205
Intervención(es)
adecuadas, 221
ensayo aleatorizado con enmascaramiento, 137
aplicación, 147 148 149 148
efectos adversos, 141 142
elección, 137 138 139
estudio preliminar, 255
sencillas, 139
únicas, 139
Introducción
de datos por duplicado, 242
al protocolo, 3
Invaginación, diagnóstico ecográfico, 190
Invención, 218
Investigación, 2 3 4 5 6 7 8 9 10 11 12 13 209
aplicada, 20 21 20
de estudios clínicos a poblaciones, 21
investigación
T1, 20
T2, 21
del laboratorio
a la cabecera, 20
a la práctica clínica, 20 21

aspectos éticos, 209 210 211 212 213 214 215 216 217 218 219 220 221
autoría, 218
autorización por el comité de revisión institucional, 211 212 211 212
comunicación de información a los participantes, 212 213
conducta científica inadecuada, 217 218
conflictos de interés, 219 220
consentimiento informado, 212 213 214 215 215
ensayos clínicos aleatorizados, 220
investigación con muestras y datos recogidos previamente, 220 221
pago a los participantes en la investigación, 221
principios éticos, 209 210
regulaciones federales, definición, 210 211
clínica.*V.* Investigación
colaborativa, 268 270 271 274
control de calidad, 262
conducta impropia, 218
control de calidad, 257 258 259 260 261 262 263
calibración, formación y certificación, 258
datos
ausentes, 261
fraudulentos, 262
gestión, 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 260 261
inexactos e imprecisos, 261 262
estudios multicéntricos colaborativos, 262
informes periódicos, 259
manual operativo, 258
práctica clínica adecuada, 257 257
procedimientos
especiales para intervenciones farmacológicas, 259
de laboratorio, 259 260
revisión del rendimiento, 258 259
cuestionarios, 223 224 225
asignación de formato, 225 226 227
determinación del marco temporal, 227 228
discordancia entre la pregunta de las opciones de respuesta, 229
enunciado, 227
escaras y puntuaciones para medir variables abstractas, 229 230
preguntas de doble efecto, 228
suposiciones ocultas, 228 229
elaboración del protocolo del estudio, 10
emparejamiento, 61
ajuste multivariante, 61 62
análisis de supervivencia, 61
estimación preliminar, 17
hipótesis
múltiples y *a posteriori,* 50 51 52
principal, 52 53
secundaria, 52 53
mediciones emparejadas, 67 68
minimización, 66 67 68 69 70
principios estadísticos, 46 47 48 49 50
prueba de la *t,* 56 57 73
tamaños de grupo desiguales, 69
variabilidad, 50 51
variables precisas, 68 69
encuestas en línea, 226 227
estudio(s)

comunitarios e internacionales, 268 269 270 271 272 273 274 275 276
aspectos
éticos, 273 274 275 274
de financiación, 272 273
barreras de distancia, idioma y cultura, 271 272
colaboración, 270 271
justificación, 268 269 270 269
recompensas, 275
riesgos y frustraciones, 275
preliminar, 255
financiación, 285 286 287 288 289 290 291
apoyo
corporativo, 289 290
interno, 290
becas
y contratos del NIH, 286 287 288 287 288
fundaciones y sociedades especializadas, 288 289
fisiología, 6 7 8 9 10
diseño del estudio, 6 7 6 10 11
ejecución del estudio, 7 8 8
errores de la investigación, 8 9 10 9 10
extracción de inferencia causal, 8
gestión de los datos, 237 238 239 240 241 242 243 244 245 246 247 248
hipótesis, 43 44 45 46
de laboratorio a la cabecera, 20
local, 269 269
mediciones, 32 33 34 35 36 37 38 39 40 41 42
escalas, 32 33 34 33
exactitud, 36 36 37 38 36 38
manual operativo, 42
muestras almacenadas, 40 41 40
otras características, 39 40
precisión, 34 35 36 38
sensibilidad y especificidad, 39
validez, 38 39
muestras sin identificación, 213 214
participativa colaborativa de base comunitaria (IPCC), 271
revisiones de protocolo, 255 256
tamaño de la muestra, 43 44 45 46 47 48 49 50 51 52 53 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82
abandonos, 60
coeficiente de correlación, 59 60 79
errores frecuentes, 71 72
estimación, 70 71
estudios
analíticos y experimentos, técnicas, 55 56 57 58 59 60 56
descriptivos, técnicas, 63 64 65
de equivalencia, 62 63
fijo, 65 66
hipótesis, 6 43 44 45 46
información insuficiente, 70 71
muestras por conglomerados, 61
prueba de la ji al cuadrado (χ^2), 57 58 59 75
variable(s)
categóricas, 60 61
continuas, 63 64 66 67
dicotómicas, 64 65

de respuesta frecuente, 69 70
Investigador(es)
clínico, 20 21 262
principal(es) (IP), 192 251 252 277
noveles, 278
IP.*V.* Investigador principal (IP)
IPCC.*V.* Investigación participativa colaborativa de base comunitaria (IPCC)
J
Juicio, 11 29 47
carácter del investigador, 211
Justicia, principio, 210
K
κ (kappa), 35 174
medición del acuerdo interobservador, 188 188
Kendall, Tau, 201
L
Laboratorio
investigador de laboratorio, 21
mediciones, 104
procedimientos, control de calidad, 259 260
Lactancia materna, uso temprano de fórmula limitada, efecto, 12 13
Lan y DeMets, método, 168
Lenguaje de búsqueda estructurado (*Structured query language,* SQL), 246
Liderazgo y formación de equipos, 252
Líderes de opinión, 289
Likert, escalas, 229
Lista(s)
de recursos de la propuesta, 281
de verificación, revisión del rendimiento, 259
Lógica de exclusión, 242
M
Magnitud del efecto normalizada, 56
Manual
Cochrane de revisiones sistemáticas, 198
operativo, 10
control de calidad, 258
elaboración de formularios, 254
estandarización de los métodos de medición, 35
Marcación telefónica aleatoria, 102
Marcadores
indirectos, variables de respuesta clínicas, 140
intermedios, variables de respuesta clínica, 140
Marihuana, 124 131
Más de una variable de respuesta, 110 149
Materiales almacenados, mediciones, 40 41 40
ME.*V.* Memorandos de entendimiento (ME)
Mecanismo
de múltiples IP, 278
seudoaleatorio, 155
Mediador, 122
Mediana, 70 71
Medias y proporciones, 63
Medición(es)
acuerdo interobservador, 188 188
adecuadas, 39
asociación, 111
cumplimiento del protocolo, 160
emparejadas, 67 68

error, 10 34 35 36 36
escalas, 32 33 34 33
estudio(s), 283
auxiliares, 196 197
de cohortes prospectivo, 88 90
transversal, 85 86
exactitud, 36 36 37 38 36 38
manual operativo, 42
minimización del sesgo, 119
muestras almacenadas, 40 41 40
precisión, 34 35 36 38
retrospectiva, 100
sensibilidad y especificidad, 39
sesgadas, 102 103 104 103
variables
abstractas, 229 230
iniciales en ensayos aleatorizados con enmascaramiento, 144 145
Medidas
de consenso, 260
existentes, 231
MediData RAVE, 246 248
MEDLINE, 198
Mejoría neta por la reclasificación (MNR), 179
Memorandos de entendimiento (ME), 271
Mentor, 16
Metaanálisis, 197 200 201 205 206
Metadatos, 240
Método(s)
estadísticos, 153
de referencia, 37 172 176 177
diagnóstico, 173
enema opaco, 190
único, aplicación selectiva, 186
Mezcla estándar, 260
MNR.*V.* Mejoría neta por la reclasificación (MNR)
Modelo
de colaboración
ascendente, 271
descendente, 271
de efectos
aleatorios, metaanálisis, 205
fijos en metaanálisis, 205
Muestra(s), 23 24 23 40
aleatoria, 28 105
estratificada, 28
simple, 28
aspectos éticos en la investigación con muestras y datos recogidos previamente, 220 221
de bancos, 40
ensayo clínico, 145
estudios auxiliares, 196
base poblacional, 27
por conglomerados, 28 61
consecutiva, 27 28 175 189
de conveniencia, 27 28
duplicados con enmascaramiento, mezclas estándar y medidas de consenso, 260
ensayo clínico, 145
estudio(s), 24 25
auxiliares, 196

de casos y testigos, 98
de cohortes prospectivo, 88
real, 24
prevista, 24
sin identificación en investigación, 213 214
probabilística, 28
representativa, 29
de sangre neonatales, investigación, 213
sistemática, 28
Muestreo, 27 28 29
emparejamiento, 124
error, 9
estudios
exactitud de pruebas, 175 176
pruebas médicas, 173
opciones de diseño, resumen, 28 29
probabilístico, 28
no probabilístico, 27 28
sesgo, 100 101 102 100
Múltiples hipótesis no relacionadas, 52
N
National Health and Nutrition Examination Survey (NHANES), 27 85 193
National Institutes of Health (NIH), 163 193 277 278 279 280 286 287 288 287 288
National Registry of Myocardial Infarction, 193 194
Neutralidad del enunciado del cuestionario, 227
NHANES.*V.* National Health and Nutrition Examination Survey (NHANES)
Niño participante en investigación, 216
Nivel
de confianza, 63
de significación estadística, 46 48
NNT.*V.* Número que es necesario tratar (NNT)
Normalización, 239
calibración, formación, certificación, 258
métodos de medición, 35
procedimiento de entrevista, 232 233
prueba de la *t,* 56 57
Novedad de la pregunta de la investigación, 18
Nueva(s)
mediciones, 196
tecnología, origen de la pregunta de la investigación, 16
Nuevos
fármacos, 158
dispositivos médicos, 291
investigadores, 278
métodos, 16
Número
adecuado, 43
de identificación de participante (ID), 237
que es necesario tratar (NNT), 71
O
O'Brien-Fleming, método, 168
Objetividad, 40
Objetivos específicos, 281 282
Observación, origen de la pregunta de la investigación, Σ 16
Obstáculos burocráticos, 275
Oracle InForm, 246
Osteoporosis, 193
P

P_1 y P_2, 58
Países de renta baja o media (PRBM), 268 272 274
Paquete estadístico, 245 248
Partición cursiva, 180
Participación local, investigación clínica, 275
Participante(s), 210 256
en estudios, 5 23 24 25 26 27 28 29 30 31 23
aspectos éticos, 209 210 211 212 213 214 215 216 217 218 219 220 221
autoría, 218
autorización por el comité de revisión institucional, 211 212 211 212
conducta científica inadecuada, 217 218
conflictos de interés, 219 220
consentimiento informado, 212 213 214 215 215
ensayos clínicos aleatorizados, 220
investigación en muestras y datos recogidos previamente, 220 221
pago a los participantes en la investigación, 221
principios éticos, 209 210
ausencia de capacidad de toma de decisiones, 215
criterios
de exclusión, 27
de inclusión, 26 27 26
cumplimiento del protocolo, 160 161 162 160
diseño del protocolo para la selección, 25
ensayo aleatorizado con enmascaramiento
asignación aleatoria a los grupos del estudio, 145 146 147 148 149
grupo testigo, elección, 139
selección, 142 143 144 143
variables iniciales, medición, 144 145
estudio
de casos y testigos, 97
de cohortes, 94 95
generalización de los hallazgos del estudio, 24 25 24
inclusión, 29 30
muestra(s)
de conveniencia, 27 28
probabilísticas, 28
número, 17
población(es)
clínicas y comunitarias, 27
y muestra objetivo, 24
resumen de las opciones de diseño de muestreo, 28 29
sección de ética de la propuesta, 286
supervisión de ensayos clínicos, 163 164 164
humanos
aspectos éticos, 284
regulaciones federales para la investigación, 210 211 212 213 214 215 216
humanos.*V.* Participantes en estudios
médico responsable del tratamiento, 216
pago por la investigación, 221
selección, 142 143 144
simulado, 255
vulnerables, protecciones, 216 217
Patrón con forma de U, 34
PCC.*V.* Práctica clínica correcta (PCC)
Perdida durante el seguimiento, 63 86 94
Período de reposo farmacológico, 157
Perjuicio, interrupción, 163

Persona-tiempo, 94
Personal
formación, 261 263
reuniones, 253
Peticiones
de propuestas (RFP), 286
de solicitudes (RFA), 286
Plagio, 218
Plan(es)
de análisis, 283
del estudio
características de una buena propuesta, 285
directrices para la supervisión, 164
elaboración de la pregunta de la investigación, 19
minimización del sesgo, 119 120 121
en una página, 10 11
Plazos, 278
Población(es), 23 24 23
accesible, 24 29 30
clínica, 27
investigadores, 22
muestra, 27 101 102
objetivo, 24 24
variadas, 21
vulnerables, 221
POE.*V.* Procedimientos operativos estándar (POE)
Posibilidad de generalización, 6 23 26
especificación, 123
estudios de pruebas médicas, 173
investigación comunitaria, 269
muestreo probabilístico, 28
Posibles
beneficios, y riesgos, 212
factores de confusión, 94 128
Potencia, 48 49 48
condicional, 168
hipótesis, 6
mediciones emparejadas, 67 68
precisión, 34 68 69
revisión sistemática, 202
tamaños de grupos desiguales, 69
variables
continuas, 66 67
de respuesta frecuente, 69 70
Práctica clínica correcta (PCC), 257 257
Precisión, 34 35 36 36 38
emparejamiento, 124
estrategias de mejora, 35 36
evaluación, 34 35
mediciones, 35 118
Pregunta(s)
abiertas, 223 224 225
cerradas, 223 224 225
detectora, 226
doble efecto, 228
estrecha, 137
estudio, 11
investigación, 2 3 14 15 16 17 18 19 20 21

análisis de datos secundarios, 195 196
características, 17 18 19 17
buena propuesta, 285
diseño del estudio, 6 7 6 10 11
elaboración de un plan del estudio, 19
interés de la investigación, 15
orígenes, 14 15 16
principal, 19
única, 19
revisión sistemática, 198
secundaria, 19
sesgo, 119
utilización de investigación local, 269
sobre el marco temporal, 227 228
ramificadas, 226 235
Preinclusión activa, 162
Presos, investigación, 216 217
Presupuesto
justificación, 281
sección de una propuesta, 280 281
Prevalencia
estudio transversal, 86 97
relativa, 86
Probabilidad previa, 51 52
Problema(s)
éticos, 209 210 211 212 213 214 215 216 217 218 219 220 221
autoría, 218
conducta científica impropia, 217 218
conflictos de interés, 219 220
datos fraudulentos, 262
ensayos clínicos aleatorizados, 220
estudios
internacionales, 273 274 275 274
de pruebas diagnósticas, 184 185
investigación con muestras y datos recogidos previamente, 220 221
pago a los participantes en la investigación, 221
pregunta de la investigación, 18 19
principios éticos, 209 210
propuestas, 284 285
regulaciones federales para la investigación con participantes humanos, 210 211 212 213 214 215 216
autorización por el comité de revisión institucional, 211 212 211 212
consentimiento informado, 212 213 214 215 215
minimización de los riesgos, 215 216
supervisión de ensayos clínicos, 163 164 164
pruebas múltiples, 168 169
seguridad, ensayo aleatorizado con enmascaramiento, 138
Procedimientos
del estudio, 283
operativos estándar (POE), 257
Proceso
iterativo, 280
de revisión, 286
Programas de base de datos de sobremesa integrados, 245
Proporción elevada, 290
Propuesta(s), 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291
características de una buena propuesta, 285
elementos, 280 281 282 283 284 285
comienzo, 280

estrategia de la investigación, 282 283 284 283
objetivos específicos, 281 282
partes
administrativas, 280 281
ética y miscelánea, 284 285
sección(es)
métodos científicos, 284
objetivos e importancia, 281 282
trabajo previo, 283
financiación, 285 286 287 288 289 290 291
apoyo
corporativo, 289 290
interno, 290
becas
y contratos del NIH, 286 287 288 287 288
de fundaciones y sociedades especializadas, 288 289
iniciadas por un instituto, 286
investigación, 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291
características de una buena propuesta, 285
elementos, 280 281 282 283 284 285
comienzo, 280
estrategia de la investigación, 282 283 284 283
objetivos científicos, 281 282
partes
administrativas, 280 281
ética y miscelánea, 284 285
sección(es)
métodos científicos, 284
objetivos de importancia, 281 282
financiación, 285 286 287 288 289 290 291
apoyo
corporativo, 289 290
interno, 290
becas
y contratos del NIH, 286 287 288 287 288
de fundaciones y sociedades especializadas, 288 289
redacción, 277 278 279 280 279
modelo, 279
Protocolo, 277
completo, 159
estructura del proyecto de investigación, 2 3
del estudio, 10
diseño para seleccionar participantes en estudios, 25 25
estructura del proyecto de investigación, 2 3
finalización, 255 256 257
resumen, 280
revisiones, 255 256
sección de importancia, 3
seguimiento y cumplimiento, 160 161 162 160
finalización, 255 256 257
resumen, 280
revisiones, 255 256
sección de importancia, 3
seguimiento y cumplimiento, 160 161 162 160
Proyecto de investigación
minimización de riesgos, 215 216
naturaleza, 212
procedimientos del estudio, 212

riesgos y beneficios, 212
Prueba(s)
diagnósticas, 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190
aspectos generales, 171 172 173
cálculo de κ para medir el acuerdo interobservador, 188
determinación de la utilidad del estudio, 171 172 173 172
errores de diseño frecuentes, 185 186 187
estudio sobre los resultados, efecto, 183 184 185
exactitud de las pruebas, estudios, 175 176 177 178 179
cocientes de verosimilitudes, 177 178 179 179
curvas de rendimiento diagnóstico, 177 177
mejora neta por la reclasificación, 179
muestreo, 175 176
riesgos absolutos, cocientes de riesgos, diferencias de riesgos y cocientes de riesgos instantáneos, 179
sensibilidad y especificidad, 177
variables
predictivas, 176
de respuesta, 176
reglas de predicción clínica, 180 181
reproducibilidad de pruebas, estudios, 173 174 175
resultados de la prueba sobre las decisiones clínicas, efecto, 181 182
viabilidad, costes y riesgos de las pruebas, estudios, 182 183
estadística, 50
bilateral, 49
estimación del tamaño de la muestra, 56
homogeneidad, 206
supervisión de los resultados intermedios, 163 164 168 169
unilateral, 49
final, 255
ji al cuadrado (χ^2), 57 58 59 75 164
pronósticas, 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190
errores de diseño, 185 186 187
exactitud de las pruebas, estudios, 175 176 177 178 179
cocientes de verosimilitud, 177 178 179 179
curvas de rendimiento diagnóstico, 177 177
determinación de la utilidad del estudio, 171 172 173 172
κ para medir el acuerdo interobservador, cálculo, 188 188
mejoría neta por la reclasificación, 179
muestreo, 175 176
problemas, 171 172 173
riesgos absolutos, cocientes de riesgos, diferencias de riesgos y cocientes de riesgos instantáneos, 179
sensibilidad y especificidad, 177
variables
predictivas, 176
de respuesta, 176
viabilidad, costes y riesgos de las pruebas, estudios, 182 183
pruebas, efecto sobre los resultados, 183 184 185
reproducibilidad de pruebas, estudios, 173 174 175
resultados de pruebas, efectos sobre decisiones clínicas, 181 182
t (prueba de la *t* de Student), 164
para dos muestras, 68 82
no emparejadas, 68
mediciones emparejadas, 68
de una muestra emparejada, 68 82
tamaño de la muestra, 56 57 73
tándem, 176
Z, 58 75

unilateral, 58
Puesta en marcha del estudio, 253 254
Puntuaciones
medición de variables abstractas, 229 230
de predisposición, 128 129 131
Q
QuesGen, 246 248
R
Razón de posibilidades, 59 93 94 93 114 115
Recogida de datos
ensayo del equipo de investigación, 255
revisión sistemática, 199 200 200
revisiones del protocolo, 255 256
Redacción de la propuesta, 277 278 279 280 279
REDCap (Research Electronic Data Capture), 245 246 248
Redes de investigación en consultas, 269
Registro(s), 193 237
certificados de defunción, 193 194
estudio de casos y testigos de base poblacional, 101 102
de llamadas, 239 246
tumores, 193
Regla de predicción clínica, 180 181
Regresión a la media, 156
Regulaciones federales para la investigación con participantes humanos, 210 211 212 213 214 215 216
autorización por el comité de revisión institucional, 211 212 211 212
consentimiento informado, 212 213 214 215 215
feto participante en la investigación, 217
minimización de riesgos, 215 216
Relación
dosis-respuesta, 131
de uno a muchos, 238
Repetición, precisión, 36
Resistencia a las alteraciones, 145
Respeto por las personas, principio, 209 210
Respuesta(s)
interactiva a la voz (RIV), 233
mutuamente excluyentes al cuestionario, 224
Resultado(s)
limítrofes, 186
negativos falsos, 46 47
«no significativo», 49
positivos falsos, 46 47
Resumen, 280
concepto, 229
de los datos, 199
estudio, 10 11
del proyecto, 280
propuesta, 279 280 279
variables, medición, 229 230
Retención de los registros de la investigación, 215
Reunión de datos, 246
Revisión
acelerada, 212 212
de la bibliografía
estimación del tamaño de la muestra, 70 71
pregunta de la investigación, 15
validación de medición abstracta, 39
externa, 219 259 286

mayores del protocolo, 263
menores del protocolo, 255 256
propuesta, 280
de propuesta, 280
protocolo, 255 256
del rendimiento, 258 259
sistemática, 15 197
análisis de subgrupos y de sensibilidad, 202 203
criterios para la inclusión y la exclusión de estudios, 199
evaluación del sesgo de publicación, 201 202 202
metaanálisis, 200 201
pregunta de la investigación, 198
recogida de datos, 199 200
RFA.*V.* Peticiones de solicitudes (RFA)
RFP.*V.* Peticiones de propuestas (RFP)
Riesgo(s)
absolutos, pruebas pronósticas, 179
conjuntos, 106
elevado, 142
mínimo para los participantes, 212
relativo, 59*V.* también Cociente de riesgos
razón de posibilidades, 111 114
para la salud, 275
RIV.*V.* Respuesta interactiva a la voz (RIV)
Rofecoxib, efectos adversos cardiacos, 217 218
S
Salud pública, consecuencias, 275
San Francisco Mammography Registry, 193
Sección(es)
abordaje, estrategia de investigación, 282
administrativa de una propuesta, 280 281
del estudio, 286
de importancia
estrategia de investigación, 282
protocolo, 3
innovaciones, estrategia de investigación, 282
métodos de una propuesta, 282 283 284 283 284
objetivos y significación de una propuesta, 281 282
preguntas y problemas de la propuesta, 285
Seguimiento, 160 161 162 160
auditoría, 262
Selección, 29 30
ensayo aleatorizado con enmascaramiento, 144
muestra representativa, consecución, 29
objetivos, 29
participantes en estudios, 29 30
Sencillez del enunciado del cuestionario, 227
Sensibilidad, 39 177 189
análisis, 202 203
Serie de casos, 87
Sesgo, 37
asociación espuria, 119 120 121 120
ausencia de respuesta, 29
conflictos de interés económicos, 219
cumplimiento del protocolo, 160 161 162 160
diferencial, 130
no diferencial, 130
enmascaramiento, 37

estudio de casos y testigos, 102 103 104 103
enmascaramiento, 147 148
error sistemático, 9
espectro, 171 172
estudio de casos y testigos, 100 101 102 103 104
medición diferencial, 102 103 104 103
muestreo, 100 101 102 100
incorporación, 173
método de referencia doble, 186
observador, 37
participante, 37
publicación, 201 202 202
recuerdo, 103
reducción, 219
verificación
diferencial, 186 187 190
estudios de pruebas diagnósticas, 186 187 189 190
voluntario, 184
Significación estadística, nivel, 46 48
Sistema(s)
de estadísticas vitales, 193
frontal, 243
de soporte, 243
Skype, 272
Sociedades profesionales, 288
Sondeo en entrevista, 233
SQL.*V. Structured query language* (SQL)
Student, prueba de la *t,* 56 57 73
Subcomité, 262
Subvenciones
«K», 286
«R», 286
Suero, 203
Sueros, 145 196
Supervisión
del ensayo clínico, 163 164 165 166 167 168 169 164
intermedia, 163 164 164 168 169
Suposiciones ocultas en cuestionario, 228 229
Supresión, 130
Supuesto mejorado, 71
Surveillance, Epidemiology, and End Results (SEER), 193
Susceptibilidad al sesgo, 100
T
Tablas y figuras, 285
Tabulaciones periódicas, 262
Tamaño de la muestra, 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82
abandonos, 60
ajuste multivariante, 61 62
análisis de supervivencia, 61
emparejamiento, 61
ensayo aleatorizado con enmascaramiento, 144
errores frecuentes, 71 72
estimación, 70 71
preliminar, 17
estrategias para la minimización, 72
estudios
analíticos y experimentos, técnicas, 55 56 57 58 59 60 56
coeficiente de correlación, 59 60 79

prueba(s)
de la ji al cuadrado (χ^2), 57 58 59 75
de la *t,* 56 57 73
descriptivos, técnicas, 63 64 65
categóricas, 60 61
variables
continuas, 63 64
dicotómicas, 64 65
de equivalencia, 62 63
de pruebas diagnósticas, 185
fijo, 65 66
hipótesis, 6 43 44 45 46
múltiples y *a posteriori,* 50 51 52
principal, 52 53
secundarias, 52 53
información insuficiente, 70 71
minimización, uso, 66 67 68 69 70
mediciones emparejadas, 67 68
tamaños del grupo desiguales, 69
variable(s)
continuas, 66 67
precisas, 68 69
de respuesta frecuente, 69 70
muestras por conglomerados, 61
planificación, 43
principios estadísticos, 46 47 48 49 50
errores de tipo I y de tipo II, 46 47
hipótesis alternativas, lados, 49 50
magnitud del efecto, 47 48
α, β y potencia, 48 49
prueba estadística, tipo, 50
valor de *p,* 49
sesgo de publicación, 201
variabilidad, 50
Tasa
de incidencia, 94
de respuesta, 29
de utilización, 194
Tendencias seculares, 156
Testigo
ausencia en diseño entre grupos, 156
elección, 139
estudio de casos y testigos, 101
anidados
casos-cohortes, 105
con densidad de incidencia
de lista de espera, 157
con placebo, 139
aspectos éticos, 220 273
procedecia en estudios de casos y testigos
consulta, 101
hospital, 101
The Foundation Center, 288
Título de la propuesta, 280
Trabajo
descripción de la propuesta, 281
en red en investigación comunitaria, 270

solicitantes, 251
Tratamiento
de los colaboradores, 274
preventivo, 138
Tuskegee, estudio, 209
U
Umbral, 34
Unidad de tiempo, 227
Utilización en la práctica real y evaluación de la eficacia, 194
V
Validez, 38 39
aparente, 39 232
de constructo, 39 232
de contenido, 39 232
cuestionario, 232
externa, 6
interna, 6
predictiva, 39 232
relacionada con el criterio, 39
Valor(es)
de corte, 173
marginales, κ, 188
de *p,* 49 50 131 206
predictivo
negativo, 177
positivo, 177
Variabilidad, 50 51
cálculo, 73
estudios de pruebas médicas, 173 174
instrumento, 34
interobservador, 173 174
intraobservador, 173
observador, 34
participante, 34
Variable(s), 5
categórica(s), 33 60 61 174
acuerdo interobservador, 174
policotómicas, 33
confusión, 61 62 122
estudios de cohortes múltiples, 93
manejo, 122 123 124 125 126 127 128 129 123 127
continuas, 33
análisis de los resultados de estudios clínicos, 164
estudios descriptivos, 63 64 80
potencia, 66 67
prueba de la *t,* 73
variabilidad interobservador, medidas, 175
dependiente, 5
dicotómicas, 33
análisis de los resultados de ensayos clínicos, 164
comparación con variables continuas, 66 67
estudio(s)
descriptivos, 64 65
de viabilidad, 183
información insuficiente, 70
prueba de la *Z,* 75
dicotómicas, 33
análisis de los resultados de un ensayo clínico, 164

comparación con variables continuas, 66 67
estudios descriptivos, 64 65 81
información insuficiente, 70
prueba de la *Z*, 75
enumeración, 230
errores frecuentes, 71 72
estratificación, 126
estudio(s)
auxiliares, 196
diseño, 7
transversal, 85
etiqueta, 241
exactitud, 36 36 37 38 36 38
independiente, 5
iniciales, 144 145
mayores, 231
medición de las variables iniciales en estudios aleatorizados con enmascaramiento, 144 145
κ para medir el acuerdo interobservador, cálculo, 188 188
nombres, 240 241
nominales, 33
numéricas, 33
discretas, 33
ordinales, 33 60 61
precisas, minimización del tamaño de la muestra, 68 69
precisión, 34
predictiva, 5
análisis de datos secundarios, 195
estudios
de casos y testigos, 98 99 100
anidados, con densidad de incidencia y casos y cohortes, 104 105 105
de cohortes, 94
prospectivo, 88 89
retrospectivo, 91
de exactitud de pruebas, 176
transversal, 85 86
hipótesis, 44
minimización del sesgo, 119
reglas de predicción clínica, 180 181
variable de confusión, 122
respuesta(s), 5
adjudicación, 162
análisis de datos secundarios, 195
clínicas, 140
continua, 56 67 145
diseño con cruzamiento, 157
episodios, 159
estudio(s)
de cohortes, 94
retrospectivo, 90
sobre el efecto de las pruebas, 183 184 185
de exactitud de pruebas, 176
transversal, 85 86
frecuente, 69 70
hipótesis, 44
medición(es)
emparejadas, 67 68
ensayo aleatorizado con enmascaramiento, 140 141 142 145
minimización del sesgo, 119

principal, 140
sesgo de publicación, 201
variables, 5 13 56
análisis de datos secundarios, 195
de confusión, 122
diseño con cruzamiento, 157
estudio
de cohortes, 94
de cohortes retrospectivo, 90
transversal, 85 86
exactitud de pruebas, estudios, 176
hipótesis, 44
medición(es)
emparejadas, 67 68
en ensayo aleatorizado con enmascaramiento, 145
minimización del sesgo, 119
prueba, efectos sobre los resultados, 184
variable de confusión, 122
resumen, 229 230
Variación regional, 194
Viabilidad
estudio, 182 183
pregunta de la investigación, 17 18 17
revisión sistemática, 198
selección de los participantes en el estudio, 29
Violación de la confidencialidad, 215
Visión general del estudio, 10 11
Visita de selección, 161 162
Vulnerabilidad
desventajas sociales y económicas, 216
deterioros cognitivos/comunicativos, 216
diferencias de poder, 216
W
Women's Health Study, 151

booksmedicos.org