NDS

Epidemiología clínica

5.ª edición

booksmedicos.org

Robert Fletcher
Suzanne W. Fletcher
Grant S. Fletcher

Wolters Kluwer

Epidemiología clínica

NDS

5.ª edición

Robert H. Fletcher, M.D., M.Sc.
Professor Emeritus
Department of Population Medicine,
Harvard Medical School,
Boston, Massachusetts

Adjunct Professor
Departments of Epidemiology and Social Medicine,
The University of North Carolina at Chapel Hill,
Chapel Hill, North Carolina

Suzanne W. Fletcher, M.D., M.Sc.
Professor Emerita
Department of Population Medicine,
Harvard Medical School,
Boston, Massachusetts

Adjunct Professor
Departments of Epidemiology and Social Medicine,
The University of North Carolina at Chapel Hill,
Chapel Hill, North Carolina

Grant S. Fletcher, M.D., M.P.H.
Assistant Professor of Medicine
The University of Washington School of Medicine,
Seattle, Washington

Wolters Kluwer
Philadelphia • Baltimore • New York • London
Buenos Aires • Hong Kong • Sydney • Tokyo

booksmedicos.org

Av. Carrilet, 3, 9.ª planta – Edificio D
Ciutat de la Justícia
08902 L'Hospitalet de Llobregat
Barcelona (España)
Tel.: 93 344 47 18
Fax: 93 344 47 16
e-mail: consultas@wolterskluwer.com

NDS

Traducción de esta 5.ª edición
Dra. Diana Jiménez González
Especialista en Ginecología y Obstetricia, Subespecialista en Medicina Materno-Fetal,
Hospital Médica Sur Lomas

Traducción de la 4.ª edición
Xavier Sales Roselló
Traductor

Olga Rafart de Nadal
Traductora

Revisión científica
Dr. Malaquías López Cervantes
Jefe del Departamento de Salud Pública,
Facultad de Medicina,
Universidad Nacional Autónoma de México

Dirección editorial: Carlos Mendoza
Editora de desarrollo: Núria Llavina
Gerente de mercadotecnia: Juan Carlos García
Composición: Servei Gràfic NJR, S.L.U.
Diseño de portada: por Sonia Bocharán
Impresión: C&C Offset-China/Printed in China

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 © 2016 Wolters Kluwer
ISBN edición en español: 978-84-16353-91-0
Depósito legal: M-40653-2015

Edición en español de la obra original en lengua inglesa *Clinical Epidemiology. The Essentials, 5th edition,* de Robert H. Fletcher, Suzanne W. Fletcher y Grant S. Fletcher, publicada por Wolters Kluwer

Copyright © 2016 Wolters Kluwer

Two Commerce Square
2001 Market Street
Philadelphia, PA 19103
ISBN edición original: 978-1-4511-4447-5

Prefacio

Este libro es para los clínicos, médicos, enfermeras, asistentes de médicos, psicólogos, veterinarios y otros dedicados a la atención de los pacientes que desean comprender por sí mismos la validez de la información en la que basan sus decisiones clínicas. Los estudiantes de epidemiología y salud pública también lo pueden encontrar útil, como un complemento a los muchos y excelentes libros de texto sobre epidemiología.

Para desarrollar todo su potencial, los médicos modernos deben tener una comprensión básica de la epidemiología clínica por muchas razones.

En primer lugar, deben tomar a diario innumerables decisiones sobre la atención de los pacientes, algunas con apuestas muy altas. Su responsabilidad es basar estas decisiones en la mejor evidencia disponible, una tarea difícil porque la evidencia es amplia y está cambiando continuamente. En el nivel más simple, la atención responsable se puede lograr siguiendo recomendaciones cuidadosamente preparadas que se encuentran en las directrices basadas en la evidencia, los artículos de revisión o los libros de texto, pero la atención médica óptima es mucho más que esto. De lo contrario, sería suficiente que la llevaran a cabo técnicos que siguieran protocolos. Algunas veces la evidencia es contradictoria, haciendo que las decisiones sean a «cara o cruz». La evidencia puede ser débil, pero aún se debe tomar una decisión. Las recomendaciones que en general son correctas para los pacientes necesitan adaptarse a la enfermedad específica, las conductas y las preferencias de pacientes individuales. Los especialistas pueden estar en desacuerdo, dejando la responsabilidad principal a los médicos poniendo al paciente en el centro. Los intereses especiales –de compañías con fines de lucro o médicos cuyo ingreso o prestigio están relacionados con su consejo– pueden afectar a la forma en que se resume la evidencia. Por estas razones, los médicos deben ser capaces de sopesar la evidencia con el fin de cumplir con sus responsabilidades como profesionales para la atención de la salud.

En segundo lugar, la epidemiología clínica es ahora una parte central de los esfuerzos mundiales para mejorar la efectividad de la atención del paciente. Los médicos comprometidos con carreras de investigación están

llevando a cabo un entrenamiento formal de posgrado en métodos de investigación, a menudo en los departamentos de epidemiología. Las becas para la investigación clínica se juzgan en gran medida de acuerdo con los principios descritos en este libro. La epidemiología clínica es el lenguaje de las revistas revisadas por colegas y el «comité de selección», que decide si los informes de investigación deben ser publicados y determina qué revisiones son necesarias para que sean adecuados. Los *journal clubs* locales ahora evalúan cuidadosamente los artículos fundamentales en lugar de explorar los contenidos de varias revistas. En la actualidad, la *National Library of Medicine* incluye términos de búsqueda en MEDLINE sobre los métodos de investigación, como el ensayo aleatorizado controlado y el metaanálisis. En resumen, la medicina clínica y la epidemiología van enfocadas a una causa común. «La curación del cisma», como Kerr White lo denominó.

En tercer lugar, la atención de los pacientes debe ser divertida. Simplemente no es divertido seguir el consejo de todos los demás sin realmente saber lo que está detrás de él. Es agotador trabajar con una literatura médica vasta, o incluso con algunas cuantas revistas semanales, sin una forma para decidir rápidamente qué artículos tienen validez científica y son clínicamente relevantes, y los que no merecen la pena. Es desconcertante tomar decisiones con grandes implicaciones sin saber muy bien si en realidad son correctas. Para aprovechar toda la diversión de su profesión, los médicos necesitan tener confianza en su capacidad para pensar sobre la evidencia por sí mismos, incluso si alguien más de antemano ha hecho el trabajo más pesado al encontrar y ordenar la evidencia por tema y calidad. Es divertido estar en condiciones de participar con seguridad en las discusiones sobre la evidencia clínica, con independencia de si está dentro de su especialidad (pues todos nosotros no somos especialistas en todo aquello fuera de nuestra especialidad).

En este libro hemos ilustrado los conceptos con ejemplos sobre la atención del paciente y la investigación clínica en lugar de usar ejemplos hipotéticos, pues la medicina se basa profundamente en decisiones prácticas sobre pacientes reales. Las preguntas y los estudios importantes han evolucionado con rapidez y hemos actualizado los ejemplos para reflejar este cambio, manteniendo los ejemplos que representan aspectos atemporales en la atención de los pacientes y los estudios clásicos.

A medida que la epidemiología clínica se consolida definitivamente

dentro de la medicina, los lectores esperan más de un libro de texto básico. Por lo tanto, hemos añadido temas nuevos en esta edición. Entre ellos están la efectividad comparativa, los ensayos de práctica clínica, los ensayos de no inferioridad, los metaanálisis individuales y los conceptos modernos para clasificar las recomendaciones basadas en la evidencia. También hemos discutido sobre el riesgo, la confusión y la modificación del efecto con mayor profundidad.

El diseño de investigación moderno y los análisis, apoyados por computadoras poderosas, permiten responder preguntas clínicas con un nivel de validez y generalización con el que no se hubiera soñado hace apenas algunos años. Sin embargo, esto es posible a menudo a costa de la complejidad, alejando a los lectores de los datos reales y su significado. Muchos de nosotros podemos confundirnos conforme los científicos altamente especializados debaten sobre el significado alternativo de términos específicos o promueven nuevos enfoques para el diseño de un estudio y los análisis estadísticos, algunos de los cuales parecen incómodos, como cajas negras, sin importar lo mucho que intentemos entrar en ellas. En estas situaciones, es especialmente valioso continuar basándonos en los conceptos básicos de la investigación clínica. Hemos tratado de hacer eso comprendiendo que los lectores bien pueden querer aprender más sobre este campo de lo que sería posible sólo con un libro de texto introductorio.

La epidemiología clínica se considera ahora una parte central de un movimiento más amplio, la medicina basada en la evidencia. Con ello se reconoce la importancia, además de juzgar la validez y generalización de los resultados de la investigación clínica, de formular preguntas que puedan responderse mediante la investigación, la búsqueda de la evidencia disponible y la aplicación de lo mejor de esta evidencia en la atención de los pacientes. Siempre hemos considerado que estas competencias adicionales son importantes, y les prestamos aún más atención en esta edición.

Esperamos que los lectores experimenten tanto placer y comprensión durante la lectura de este libro como las que tuvimos nosotros al escribirlo.

Robert H. Fletcher
Suzanne W. Fletcher
Grant S. Fletcher

Agradecimientos

Tenemos la suerte de haber aprendido la epidemiología clínica de sus fundadores. Kerr White fue el tutor de Bob y Suzanne durante sus estudios de posgrado en Johns Hopkins y los convenció que lo importante son «los beneficios de las intervenciones médicas en relación con sus riesgos y costes». Alvan Feinstein enseñó a una generación de jóvenes clínicos estudiosos sobre la «arquitectura de la investigación clínica» y la dignidad de la erudición clínica. Archie Cochrane pasó una noche en nuestra casa en Montreal, cuando Grant era un niño, y nos abrió los ojos a la «efectividad y la eficiencia». David Sackett afirmó que la epidemiología clínica es una «ciencia básica para la medicina clínica» y ayudó a que el mundo la comprendiera. Muchos otros lo han seguido. Estamos especialmente agradecidos por nuestro trabajo en común Brian Haynes, editor fundador del *ACP Journal Club*; Ian Chalmers, quien hizo posible la Cochrane Collaboration; Andy Oxman, director de los talleres de atención para la salud basada en la evidencia de Rocky Mountain; Peter Tugwell, líder fundador de la *International Clinical Epidemiology Network* (INCLEN), y a Russ Harris, nuestro colega desde hace mucho tiempo en la frontera entre la medicina clínica y la salud pública en la University of North Carolina. Estas personas extraordinarias y sus colegas han creado un ambiente intelectual emocionante que dio lugar a una revolución en los estudios clínicos, llevando a un nuevo nivel a la medicina clínica basada en la evidencia.

Como todos los maestros, también hemos aprendido de nuestros estudiantes, médicos de todas las edades y de todas las especialidades que querían aprender a juzgar la validez de las observaciones clínicas y la investigación por sus propios medios. Bob y Suzanne agradecen a los estudiantes de medicina de la McGill University (quienes sugirieron por primera vez la necesidad de este libro), a la University of North Carolina y a la Harvard Medical School, a los becarios del Robert Wood Johnson Clinical Scholars, la INCLEN y la Harvard General Medicine Fellowship, a los becarios CRN del Cancer Research Network, un consorcio de institutos de investigación en sistemas de salud integrados, y a los participantes de los talleres de atención para la salud basada en la evidencia de Rocky Mountain.

Ellos fueron nuestros estudiantes y ahora son nuestros colegas; muchos enseñan e investigan con nosotros. Con los años, Grant ha conocido a muchas de estas personas y ahora aprende de estudiantes médicos, residentes y colegas académicos, conforme les enseña sobre la atención de los pacientes en el Harborview Hospital y la University of Washington.

Mientras eran editores de las revistas *Journal of General Internal Medicine* y *Annals of Internal Medicine,* Bob y Suzanne aprendieron de sus compañeros editores, incluyendo los miembros de la World Association of Medical Editors (WAME), a elaborar informes de investigación de una manera más completa, clara y equilibrada para que los lectores puedan comprender el mensaje con el menor esfuerzo. Con nuestros colegas de UpToDate, una fuente electrónica de información para médicos y pacientes, hemos desarrollado nuevas maneras de hacer que la mejor evidencia esté disponible en el mundo real, preguntas clínicas de fácil acceso durante la atención de los pacientes y facilitar la comprensión de la evidencia no sólo para los académicos y los investigadores, sino también para los médicos de tiempo completo.

Ed Wagner estaba con nosotros al comienzo de este proyecto. Con él desarrollamos un nuevo curso de epidemiología clínica para la University of North Carolina School of Medicine, y además escribió la 1.ª edición de este libro para este propósito. Después, el curso fue la introducción de Grant en la materia, en un grado en el que aún no se había introducido. Ed continuó siendo coautor en las tres ediciones y después se trasladó a la dirección del Group Health Research Institute, además de tener otras responsabilidades en Seattle. Afortunadamente, Grant está ahora en el equipo de redacción y contribuyó con su experiencia en la aplicación de la epidemiología clínica y la práctica actual de la medicina, especialmente en la atención de pacientes muy enfermos.

Damos las gracias a los miembros del equipo dirigido por Lippincott Williams & Wilkins, quienes hicieron la traducción del texto y dibujaron las figuras a mano, convirtiendo el libro en un texto moderno y atractivo. Obtuvimos una atención experta y personal de Catherine Noonan, quien nos guió en la preparación a lo largo de este libro; Jonathan Dimes, quien trabajó estrechamente con nosotros preparando las ilustraciones, y Jeri Litteral, que colaboró con nosotros en la fase de revisión editorial de este proyecto.

Agradecemos especialmente a los lectores de todo el mundo por sus comentarios alentadores y sugerencias prácticas. Nos han apoyado en las

dificultades para preparar la 5.ª edición de un libro de texto publicado por primera vez hace 30 años.

Índice de capítulos

NDS

1. Introducción
2. Frecuencia
3. Anormalidad
4. Riesgo: principios básicos
5. Riesgo: exposición a la enfermedad
6. Riesgo: de la enfermedad a la exposición
7. Pronóstico
8. Diagnóstico
9. Tratamiento
10. Prevención
11. Azar
12. Causa
13. Resumen de la evidencia
14. Administración del conocimiento

Apéndice A Respuestas a las preguntas para la revisión

Apéndice B Lecturas adicionales

Índice alfabético de materias

booksmedicos.org

Contenido

CAPÍTULO 1: INTRODUCCIÓN

Preguntas clínicas y epidemiología clínica
Resultados sobre la salud

Bases científicas para la medicina clínica

Principios básicos
Variables
Números y probabilidad
Poblaciones y muestras
Sesgo (error sistemático)
- *Sesgo de selección*
- *Sesgo de medición*
- *Confusión*

Azar
Los efectos del sesgo y del azar son acumulativos
Validez interna y externa

Información y decisiones

Organización de este libro

CAPÍTULO 2: FRECUENCIA

¿Pueden las palabras sustituir adecuadamente a los números?

Prevalencia e incidencia
Prevalencia
Incidencia
Prevalencia e incidencia con respecto al tiempo

booksmedicos.org

Relaciones entre prevalencia, incidencia y duración de la enfermedad

Otras tasas

Estudios de prevalencia e incidencia
Estudios de prevalencia
Estudios de incidencia
Incidencia acumulada
Densidad de incidencia (años-persona)

Elementos básicos de las medidas de frecuencia
¿Qué es un caso?: definición de numerador
¿Qué es la población?: definición de denominador
¿La muestra del estudio representa a la población?

Distribución de la enfermedad según el tiempo, el lugar y las personas
Tiempo
Lugar
Personas

Usos de los estudios de prevalencia
¿Cuál es la utilidad de los estudios de prevalencia?
¿Para qué no son particularmente buenos los estudios de prevalencia?

CAPÍTULO 3: ANORMALIDAD

Tipos de datos
Datos nominales
Datos ordinales
Datos de intervalo

Rendimiento de las mediciones
Validez
- *Validez del contenido*
- *Validez de criterio*
- *Validez del constructo*

Fiabilidad

Amplitud
Sensibilidad
Interpretabilidad

Variación
Variación derivada de la medición
Variación derivada de las diferencias biológicas
Variación total
Efectos de la variación

Distribuciones
Descripción de las distribuciones
Distribuciones reales
Distribución normal

Criterios de anormalidad
Anormal = Infrecuente
Anormal = Asociado a enfermedad
Anormal = Tratar el trastorno conduce a un mejor resultado clínico

Regresión a la media

CAPÍTULO 4: RIESGO: PRINCIPIOS BÁSICOS

Factores de riesgo

Reconocer el riesgo
Latencia prolongada
Causas inmediatas frente a distantes
Exposición habitual a factores de riesgo
Baja incidencia de enfermedad
Riesgo pequeño
Múltiples causas y múltiples efectos
Los factores de riesgo pueden ser o no causales

Predicción del riesgo
Combinación de múltiples factores de riesgo para predecir el riesgo

booksmedicos.org

Predicción del riesgo en pacientes individuales y grupos

Evaluación de las herramientas de predicción de riesgo
Calibración
Discriminación
Sensibilidad y especificidad de una herramienta de predicción del riesgo
Estratificación del riesgo
¿Por qué las herramientas de predicción del riesgo no discriminan bien entre los individuos?

Usos clínicos de los factores de riesgo y las herramientas de predicción del riesgo
Factores de riesgo y probabilidad preprueba de enfermedad en las pruebas diagnósticas
Uso de los factores de riesgo para elegir un tratamiento
Estratificación del riesgo en los programas de tamizaje
Eliminar factores de riesgo para prevenir la enfermedad

CAPÍTULO 5: RIESGO: EXPOSICIÓN A LA ENFERMEDAD

Estudios de riesgo
Cuando no es posible realizar experimentos
Cohortes
Estudios de cohorte
Estudios de cohorte prospectivos y de cohorte históricos
Estudios de cohorte prospectivos
Estudios de cohorte históricos utilizando bases de datos clínicas
Estudios de caso-cohorte
Ventajas e inconvenientes de los estudios de cohorte

Formas de expresar y comparar el riesgo
Riesgo absoluto
Riesgo atribuible
Riesgo relativo

booksmedicos.org

Interpretación del riesgo atribuible y el riesgo relativo
Riesgo de la población

Tener en cuenta otras variables
Variables extrañas
Descripciones simples sobre el riesgo

Confusión
Definición operativa
Posibles factores de confusión
Confirmación de la confusión

Control de la confusión
Aleatorización
Restricción
Pareamiento
Estratificación
Estandarización
Ajuste multivariado
Estrategia general para controlar la confusión

Estudios observacionales y causa

Modificación del efecto

CAPÍTULO 6: RIESGO: DE LA ENFERMEDAD A LA EXPOSICIÓN

Estudios de casos y controles

Diseño de los estudios de casos y controles
Selección de casos
Selección de controles
El método poblacional
El método de cohorte
Controles hospitalarios y de la comunidad
Múltiples grupos de control

Múltiples controles por caso
Pareamiento
Medida de la exposición
Exposiciones múltiples

Razón de momios *(odds ratio)*: una estimación del riesgo relativo

Control de las variables extrañas

Investigación de un brote epidémico

CAPÍTULO 7: PRONÓSTICO

Diferencias entre los factores pronósticos y de riesgo
Los pacientes son diferentes
Los resultados son diferentes
Las tasas son diferentes
Los factores pueden ser diferentes

Curso clínico y evolución natural de la enfermedad

Elementos de los estudios sobre pronóstico
La muestra de pacientes
Tiempo cero
Seguimiento
Resultados de la enfermedad

Descripción del pronóstico
Balance: sencillez frente a mayor información
Análisis de supervivencia
Supervivencia de una cohorte
Curvas de supervivencia
Interpretación de las curvas de supervivencia

Identificación de factores pronósticos

Serie de casos

booksmedicos.org

Reglas de predicción

Sesgo en los estudios de cohortes
Sesgo de muestreo
Sesgo de migración
Sesgo de medición
Sesgos por una clasificación errónea no diferencial

¿Es realmente importante el sesgo?

Análisis de sensibilidad

CAPÍTULO 8: DIAGNÓSTICO

Simplificación de los datos

Exactitud del resultado de una prueba
El estándar de oro
Falta de información sobre las pruebas negativas
Falta de información sobre los resultados de las pruebas en pacientes sanos
Falta de patrones objetivos para las enfermedades
Consecuencias de los patrones de referencia imperfectos

Sensibilidad y especificidad
Definiciones
Utilización de pruebas sensibles
Utilización de pruebas específicas
Equilibrio entre sensibilidad y especificidad
La curva ROC

Establecer la sensibilidad y la especificidad
Espectro de pacientes
Sesgos
Azar

Valor predictivo
Definiciones

booksmedicos.org

Determinantes del valor predictivo
Estimación de la prevalencia (probabilidad preprueba)
Aumentando la probabilidad preprueba de la enfermedad
Características específicas de la situación clínica
Grupos demográficos seleccionados
Proceso de derivación
Implicaciones para la interpretación de la bibliografía médica

Razones de verosimilitud
Momios o posibilidades *(odds)*
Definiciones
Uso de las razones de verosimilitud
¿Por qué usar las razones de verosimilitud?
Cálculo de las razones de verosimilitud

Múltiples pruebas
Pruebas en paralelo
Reglas de predicción clínica
Pruebas en serie
Razones de verosimilitud en serie
Supuesto de la independencia

CAPÍTULO 9: TRATAMIENTO

Ideas y evidencia
Ideas
Evaluar ideas

Estudios sobre los efectos de los tratamientos
Estudios observacionales y experimentales de los efectos del tratamiento

Ensayos aleatorizados y controlados
Ética
Muestreo
Intervención
Grupos de comparación

booksmedicos.org

Asignación del tratamiento
Diferencias tras la aleatorización
Pacientes sin la enfermedad en estudio
Cumplimiento
Estudios o diseños con grupos cruzados
Intervenciones conjuntas
Estudio ciego o a ciegas
Valoración de los resultados

Eficacia y efectividad
Intención a tratar y ensayos explicativos

Superioridad, equivalencia y no inferioridad

Variaciones en los ensayos aleatorizados básicos

Adaptando los resultados de los ensayos a pacientes concretos
Subgrupos
Efectividad en pacientes individuales
Ensayos de *N* = 1

Alternativas a los ensayos aleatorizados
Limitaciones de los ensayos aleatorizados
Estudios observacionales de intervenciones
Bases de datos clínicas
¿Estudios aleatorizados frente a observacionales?

Fases de los ensayos clínicos

CAPÍTULO 10: PREVENCIÓN

Actividades preventivas en los ámbitos clínicos
Tipos de prevención clínica
Vacunación
Tamizaje
Orientación conductual (cambios del estilo de vida)
Quimioprevención

booksmedicos.org

Niveles de prevención
Prevención primaria
Prevención secundaria
Prevención terciaria
Confusión sobre la prevención primaria, secundaria y terciaria

Enfoque científico de la prevención clínica

Magnitud del sufrimiento

Efectividad del tratamiento
Tratamiento en la prevención primaria
- *Ensayos aleatorizados*
- *Estudios observacionales*
- *Seguridad*
- *Asesoramiento*

Tratamiento en la prevención secundaria
Tratamiento en la prevención terciaria

Problemas metodológicos al evaluar los programas de tamizaje
Tamizajes de incidencia y de prevalencia
Sesgos especiales
- *Sesgo de anticipación diagnóstica*
- *Sesgo de duración*
- *Sesgo de cumplimiento*

Criterios para realizar las pruebas de tamizaje
Sensibilidad y especificidad elevadas
Métodos de detección e incidencia para calcular la sensibilidad
Valor predictivo positivo bajo
Simplicidad y bajo costo
Seguridad
Aceptable para médicos y pacientes

Posibles efectos adversos del tamizaje
Riesgo de un resultado falso positivo
Riesgo negativo de etiquetado

booksmedicos.org

Riesgo de seudoenfermedad (sobrediagnóstico) en el tamizaje del cáncer
Incidentalomas

Cambios en las pruebas de tamizaje y tratamientos a lo largo del tiempo

Ponderar los beneficios contra los perjuicios de la prevención

CAPÍTULO 11: AZAR

Dos enfoques del azar

Verificación de la hipótesis
Resultados falsos negativos y falsos positivos
Concluyendo sobre el buen funcionamiento de un tratamiento
Valores de *P* dicotómicos y exactos
Significación estadística e importancia clínica
Pruebas estadísticas
Conclusión sobre el mal funcionamiento de un tratamiento

¿Cuántos pacientes se necesitan para un estudio?
Potencia estadística
Requisitos para calcular el tamaño de la muestra
- *Magnitud del efecto*
- *Error de tipo I*
- *Error de tipo II*
- *Características de los datos*
- *Interrelaciones*

Un tamaño muestral pequeño fue adecuado
Un tamaño muestral grande fue inadecuado

Estimaciones puntuales e intervalos de confianza
Potencia estadística tras finalizar el estudio

Detección de episodios raros

Comparaciones múltiples

Análisis de subgrupos

booksmedicos.org

Resultados múltiples

Métodos multivariados

Razonamiento bayesiano

CAPÍTULO 12: CAUSA

Principios básicos
Causas simples
Múltiples causas
Proximidad de causa-efecto

Evidencia indirecta de causalidad
Examinando los estudios individuales
Jerarquía de los diseños de investigación

El cuerpo de la evidencia a favor y en contra de la causa
¿La causa precede al efecto?
Solidez de la asociación
Relaciones dosis-respuesta
Asociaciones reversibles
Consistencia
Plausibilidad biológica
Especificidad
Analogía

Estudios de riesgo agregado

Modelado

Evaluando la evidencia

CAPÍTULO 13: RESUMEN DE LA EVIDENCIA

Revisiones tradicionales

Revisiones sistemáticas
Definir una pregunta específica
Encontrar todos los estudios relevantes
Limitar las revisiones a estudios científicamente sólidos y clínicamente relevantes
¿Son los estudios publicados una muestra sesgada de toda la investigación concluida?
¿Cuál es la calidad de los mejores estudios?
¿La calidad científica está relacionada con los resultados de la investigación?
Resumir los resultados

Combinación de estudios en metaanálisis
¿Son los estudios lo suficientemente similares como para justificar la combinación?
¿Qué son estudios o pacientes combinados?
¿Cómo se combinan los resultados?
Identificar las causas de heterogeneidad

Metaanálisis acumulativos

Revisiones sistemáticas de estudios observacionales y pruebas diagnósticas

Fortalezas y debilidades de los metaanálisis

CAPÍTULO 14: ADMINISTRACIÓN DEL CONOCIMIENTO

Principios básicos
¿Lo hace uno mismo o delega el trabajo?
¿Qué tipo de recursos?
Clasificación de la información
Informes engañosos sobre los hallazgos de investigación

Búsqueda de respuestas a cuestiones clínicas
Soluciones

booksmedicos.org

Colegas clínicos
Libros de texto electrónicos
Guías de práctica clínica
La biblioteca Cochrane
Bases de datos (PubMed y otras)
Otros recursos en internet

Observación de progresos

Revistas
«Lectura» de revistas

Guiar a los pacientes para buscar información sobre salud

Poner en práctica la administración de los conocimientos

APÉNDICE A – RESPUESTAS A LAS PREGUNTAS PARA LA REVISIÓN

APÉNDICE B – LECTURAS ADICIONALES

ÍNDICE ALFABÉTICO DE MATERIAS

booksmedicos.org

Capítulo 1

Introducción

Debemos estudiar «los beneficios de las intervenciones médicas en relación con sus riesgos y costos».

—Kerr L. White
1992

PALABRAS CLAVE

Epidemiología clínica
Ciencias clínicas
Ciencias demográficas
Epidemiología
Medicina basada en la evidencia
Investigación de servicios de salud
Toma de decisiones cuantitativa
Análisis de costoefectividad
Análisis de decisión
Ciencias sociales
Ciencias biológicas
Variables
Variable independiente
Variable dependiente
Variables extrañas
Covariables
Poblaciones
Muestra
Inferencia

booksmedicos.org

Sesgo
Sesgo de selección
Sesgo de medición
Confusión
Azar
Variación aleatoria
Validez interna
Validez externa
Generalización
Toma de decisiones compartida

Ejemplo

Un varón de 51 años de edad solicita su asistencia a causa de un dolor torácico que cree que es «indigestión». Se encontraba bien hasta hace 2 semanas, cuando notó una opresión en el centro del pecho mientras caminaba cuesta arriba. La opresión cesó al cabo de 2-3 min de descanso. Desde entonces, se han producido molestias similares, unas veces durante el ejercicio y otras estando en reposo. Dejó de fumar un paquete de cigarrillos al día hace 3 años y se le ha comentado que su presión arterial es «un poco alta». Por lo demás, se siente bien y no toma medicamentos, si bien está preocupado por su salud, especialmente por si sufre alguna enfermedad cardíaca. Perdió su trabajo hace 6 meses y no tiene seguro médico. La exploración física completa y el electrocardiograma en reposo son normales, excepto la presión arterial, que es de 150/96 mm Hg.

Es probable que este paciente se haga muchas preguntas. ¿Estoy enfermo? ¿Con qué seguridad lo afirma? Si estoy enfermo, ¿cuál es la causa de mi enfermedad? ¿Cómo me afectará? ¿Qué debo hacer? ¿Cuánto me va a costar económicamente?

Como médico que atiende a este paciente, usted se hace el mismo tipo de preguntas, pero las suyas reflejan una mayor comprensión de sus posibilidades. ¿Es lo bastante elevada la probabilidad de que exista una enfermedad grave, pero tratable, como para ir de inmediato más allá de dar una simple explicación y de tratar de tranquilizar al paciente, y pasar a las pruebas diagnósticas? Hasta qué punto las diferentes pruebas distinguen entre las posibles causas del dolor torácico: angina de pecho, espasmo esofágico, distensión muscular, ansiedad, etc. Por ejemplo, ¿qué grado de utilidad tiene una prueba de esfuerzo para confirmar o descartar una arteriopatía coronaria? Si se detecta una enfermedad coronaria, ¿cuánto tiempo puede esperar el paciente que dure el dolor? ¿Qué posibilidades hay de que aparezcan otras complicaciones, como insuficiencia cardíaca congestiva, infarto de miocardio

booksmedicos.org

o enfermedad aterosclerótica de otros órganos? ¿La enfermedad le acortará la vida? ¿Con la disminución de los factores de riesgo de enfermedad coronaria (fumar cigarrillos e hipertensión), se reducirá su riesgo? ¿Deberían buscarse otros posibles factores de riesgo? Si los fármacos mantienen el dolor controlado, ¿podría una intervención de revascularización coronaria aumentar los beneficios, por ejemplo, evitando futuros infartos o una muerte de etiología cardiovascular? Dado que el paciente no tiene empleo ni seguro médico, ¿pueden otros estudios diagnósticos y tratamientos menos costosos lograr el mismo resultado que los más caros?

Preguntas clínicas y epidemiología clínica

Las preguntas que afrontan el paciente y el médico en el ejemplo son el tipo de preguntas clínicas que se tratan en la mayoría de los encuentros entre ambos: ¿Qué es «anormal»?, ¿Qué tan precisas son las pruebas diagnósticas que utilizamos?, ¿Cuáles son los riesgos de una enfermedad dada y cómo podemos determinarlos?, ¿La enfermedad en general empeorará, permanecerá igual o se resolverá (pronóstico)?, ¿El tratamiento realmente mejora al paciente o sólo mejora los resultados de la prueba?, ¿Existe alguna manera de prevenir la enfermedad?, ¿Cuál es la causa subyacente de la enfermedad o el trastorno?, ¿Cómo podemos ofrecer una mejor atención médica de forma más eficiente? Estas preguntas clínicas (tabla 1-1) y los métodos epidemiológicos para responderlas son la piedra angular de este libro. Cada una es también el tema de los capítulos específicos que se desarrollarán en esta obra.

Los médicos necesitan las mejores respuestas posibles a estas preguntas. Para ello, recurren a diferentes fuentes de información: su propia experiencia, el consejo de sus compañeros y el razonamiento a partir de sus conocimientos sobre la biología de la enfermedad. En muchas situaciones, la fuente más creíble es la investigación clínica, que implica utilizar observaciones anteriores de pacientes similares para predecir qué le sucederá al paciente actual. La forma en que se realizan y se interpretan dichas observaciones determina la validez de las conclusiones alcanzadas y, por lo tanto, el grado de utilidad de las conclusiones para los pacientes.

Resultados sobre la salud

Los acontecimientos más importantes en la medicina clínica son los

booksmedicos.org

resultados sobre la salud de los pacientes, como los síntomas (incomodidad y/o insatisfacción), discapacidad, enfermedad y muerte. Estos resultados enfocados en el paciente se refieren a veces como las «5 D», es decir, *Death*, *Disease*, *Discomfort*, *Disability* y *Dissatisfaction* (tabla 1-2). Son eventos sobre la salud que preocupan a los pacientes. Los médicos deben tratar de entender, predecir, interpretar y cambiar estos resultados al tratar a los pacientes. Las 5 D pueden estudiarse sólo en humanos intactos y no en partes humanas (p. ej., transmisores humorales, cultivos tisulares, membranas celulares y secuencias genéticas) o en animales. La epidemiología clínica es la ciencia que se emplea para estudiar las 5D en humanos intactos.

Tabla 1-1

Temas y preguntas clínicas[a]

Tema	Pregunta
Frecuencia (cap. 2)	¿Con qué frecuencia se produce una enfermedad?
Anormalidad (cap. 3)	¿El paciente está sano o enfermo?
Riesgo (caps. 5 y 6)	¿Qué factores se asocian a un aumento del riesgo de una enfermedad?
Pronóstico (cap. 7)	¿Cuáles son las consecuencias de sufrir una enfermedad?
Diagnóstico (cap. 8)	¿Qué grado de exactitud ofrecen las pruebas usadas para el diagnóstico de la enfermedad?
Tratamiento (cap. 9)	¿En qué medida puede cambiar un tratamiento el curso de la enfermedad?
Prevención (cap. 10)	¿Una intervención sobre las personas sanas impide que surja la enfermedad? ¿La detección y el tratamiento precoces mejoran la evolución de la enfermedad?
Causa (cap. 12)	¿Qué condiciones conducen a la enfermedad? ¿Cuáles son los orígenes de la enfermedad?

a El libro incluye cuatro capítulos: 4, 11, 13 y 14 que se relacionan con todos estos temas.

En la medicina clínica moderna, con tanto solicitar y tratar los resultados

booksmedicos.org

de laboratorio (para cosas como los niveles plasmáticos de glucosa, hematuria, troponinas, etc.), es difícil recordar que los resultados de laboratorio no son los acontecimientos (eventos) importantes para la propia práctica médica. Es fácil asumir que si podemos cambiar los resultados anómalos de laboratorio a la normalidad, se ayuda al paciente. Esto es cierto sólo en la medida en que un estudio cuidadoso haya demostrado un vínculo entre los resultados de laboratorio y una de las 5 D.

Tabla 1-2

Resultados de la enfermedad[a]	
Muerte	Un mal resultado si llegas antes de tiempo
Enfermedad[b]	Un conjunto de síntomas, signos físicos y alteraciones de laboratorio
Malestar	Síntomas como dolor, náuseas, disnea, prurito y acúfenos
Incapacidad	Deterioro de la capacidad de llevar a cabo las actividades cotidianas en casa, el trabajo o en el tiempo de ocio
Insatisfacción	Reacción emocional hacia la enfermedad y hacia sus cuidados, como tristeza o ira

[a]Quizá habría que añadir un sexto resultado a esta lista, el de pobreza, ya que los costos económicos de la enfermedad (para el paciente o la sociedad) son una consecuencia importante de la enfermedad.
[b]O bien padecimiento, la experiencia que el paciente tiene de la enfermedad.

Ejemplo

La incidencia de la diabetes mellitus tipo 2 está aumentando de forma dramática en EE.UU. El riesgo de los diabéticos de morir por una enfermedad cardíaca es dos a cuatro veces mayor que en las personas sin diabetes, y la enfermedad cardiovascular representa un 70 % de todas las muertes en pacientes diabéticos. Los recientes esfuerzos farmacológicos para controlar la diabetes han producido una clase de fármacos, las tiazolidinedionas, que aumentan la sensibilidad de la insulina en el músculo, la grasa y el hígado. Varios estudios han demostrado que estos fármacos disminuyen los niveles de hemoglobina glucosilada (HbA1C) en pacientes diabéticos. Uno de estos fármacos, la rosiglitazona, fue aprobada para su uso en 1999. Sin embargo, durante los años siguientes, varios estudios de seguimiento demostraron un resultado sorprendente: Los pacientes tratados con el fármaco eran propensos a experimentar más, no menos, problemas cardíacos, pues diversos estudios mostraron un incremento en los ataques cardíacos, insuficiencia cardíaca,

booksmedicos.org

accidente cerebrovascular o mortalidad por cualquier causa (1-3). Dado que muchos de los estudios que demostraron resultados positivos del fármaco en los niveles de glucosa y HbA1C no estaban diseñados originalmente para examinar los resultados cardiovasculares a largo plazo, la mayoría de los estudios de seguimiento no eran ensayos rigurosos. Sin embargo, se despertó la preocupación suficiente, pues en 2010 la Food and Drug Administration (FDA) restringió el uso de rosiglitazona; en Europa, su venta se suspendió.

Durante su entrenamiento, los médicos están inmersos en la biología de la enfermedad, la secuencia de pasos desde un acontecimiento subcelular hasta la enfermedad y sus consecuencias. Por lo tanto, parece razonable asumir que una intervención que disminuye la glucosa sérica en los diabéticos ayudará a protegerlos contra la enfermedad cardíaca. Sin embargo, aunque es muy importante para la medicina clínica, estos mecanismos biológicos no pueden ser sustituidos por los resultados del paciente a menos que una evidencia sólida confirme que los dos están relacionados. (De hecho, los resultados de los estudios con varios medicamentos diferentes están aumentando la posibilidad de que en la diabetes tipo 2 la disminución agresiva de los niveles de glucosa sérica no protege de la enfermedad cardíaca). En el caso de los medicamentos nuevos, es especialmente importante establecer la mejoría en los resultados sobre la salud de los pacientes porque las intervenciones farmacológicas suelen tener varios efectos clínicos en lugar de uno solo.

BASES CIENTÍFICAS PARA LA MEDICINA CLÍNICA

La epidemiología clínica es una de las ciencias básicas sobre la que se apoyan los médicos para la asistencia a los pacientes, aunque también existen otras ciencias que son parte integrante de esta asistencia, y que se resumen en la figura 1-1. Muchas de estas ciencias se superponen entre sí.

La **epidemiología clínica** es la ciencia de efectuar predicciones en pacientes individuales al considerar acontecimientos clínicos (las 5 D) en grupos de pacientes similares y el uso de métodos científicos sólidos para asegurar que las predicciones son precisas. Su propósito es desarrollar y aplicar los métodos de observación clínica que conducirán a conclusiones válidas al evitar ser engañados por el error sistemático y la intervención del azar. Es un abordaje importante para obtener el tipo de información que los clínicos necesitan para tomar buenas decisiones en la atención de los pacientes.

El término «epidemiología clínica» deriva de dos disciplinas afines: la

booksmedicos.org

medicina clínica y la epidemiología. Es «clínica» porque trata de encontrar respuesta a las preguntas clínicas y guiar, al mismo tiempo, la toma de decisiones clínicas con las mejores pruebas disponibles. Es «epidemiología» porque muchos de los métodos utilizados para responder a esas preguntas sobre cómo atender mejor a los pacientes han sido creados por epidemiólogos, y porque la asistencia a pacientes concretos se entiende en el contexto de una población de mayor tamaño de la que el paciente forma parte.

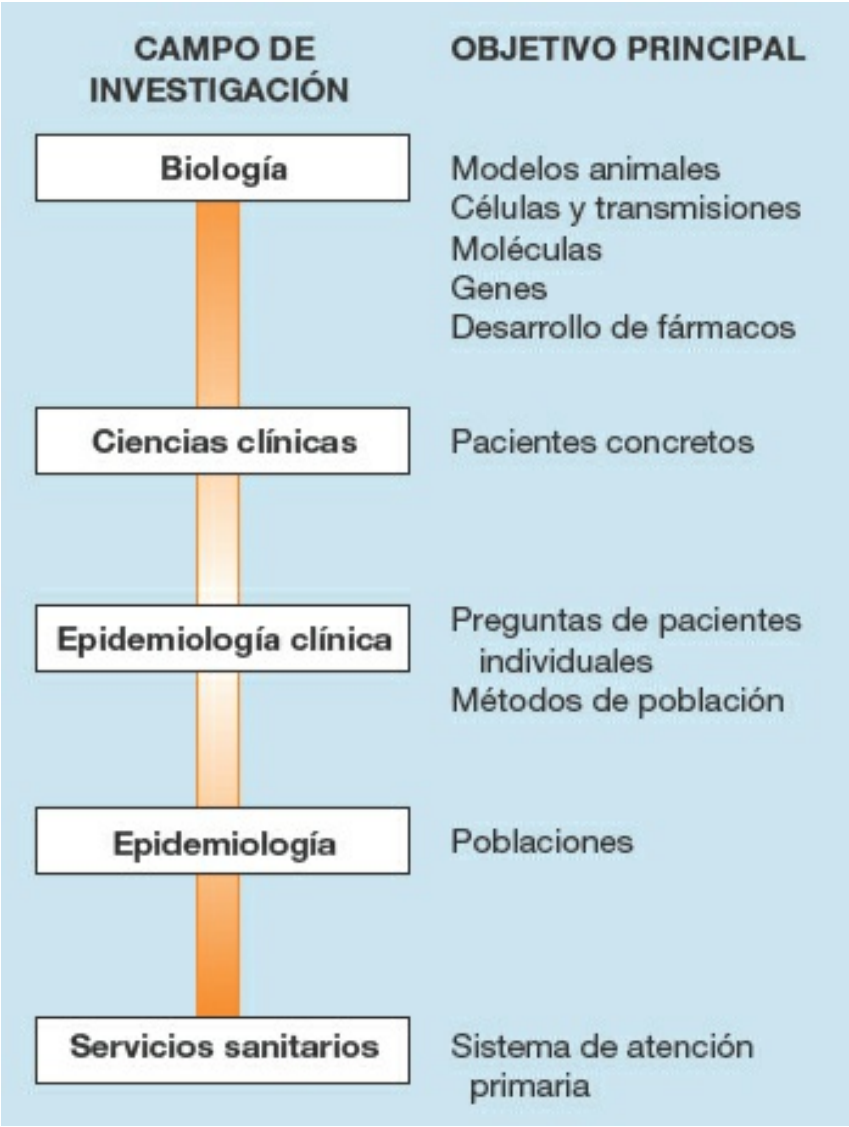


Figura 1-1 ■ Las ciencias de la salud y sus relaciones complementarias.

Las **ciencias clínicas** proporcionan las preguntas y el método de abordaje

booksmedicos.org

que pueden ser utilizados en la atención del paciente individual. Algunas ciencias biológicas, como la anatomía y la fisiología, son «clínicas» en la medida en que ofrecen información sólida sobre la que basar las decisiones clínicas. Por ejemplo, el conocimiento de la anatomía del cuerpo ayuda a determinar posibilidades para el diagnóstico y tratamiento de muchos síntomas.

Las **ciencias demográficas** estudian grandes grupos de personas. La **epidemiología** es el «estudio de la aparición de enfermedades en poblaciones humanas» (4), a partir del recuento de los episodios relacionados con la salud que presentan las personas en relación con los grupos que existen de forma natural (poblaciones) y a los cuales pertenecen dichas personas. Los resultados de muchos de estos estudios pueden aplicarse directamente a la asistencia de pacientes concretos. Por ejemplo, estos estudios se utilizan como base para aconsejar que deben evitarse comportamientos como el tabaquismo y la inactividad, ya que aumentan el riesgo de los pacientes. Otros estudios epidemiológicos, como los que muestran los efectos nocivos del tabaquismo pasivo y de otros riesgos ambientales y profesionales, sirven de base para las recomendaciones médicas. La epidemiología clínica es un subgrupo de las ciencias demográficas, útil en la atención de los pacientes.

Durante mucho tiempo, los médicos han dependido en cierto modo de las evidencias de investigación. Pero comprender las evidencias clínicas es más importante hoy de lo que fue en el pasado por varios motivos. Se dispone de una cantidad de información extraordinaria. Las intervenciones diagnósticas y terapéuticas pueden ser muy eficaces, aunque a su vez entrañan riesgos y costos, por lo que decidirse por una u otra opción puede tener sus consecuencias. La investigación clínica es, en el mejor de los casos, mucho más potente y, por lo tanto, representa una base más sólida para la toma de decisiones clínicas. La credibilidad de la investigación clínica varía entre un estudio y otro, por lo que los médicos deben contar con un método para distinguir las pruebas sólidas de las que no lo son.

La **medicina basada en la evidencia** es un término moderno que se refiere a la aplicación de la epidemiología clínica a la asistencia a los pacientes. Implica formular preguntas clínicas concretas que «puedan ser respondidas», encontrar la mejor evidencia de la investigación disponible teniendo en cuenta dichas preguntas, juzgar la evidencia por su validez e integrar la evaluación crítica con la experiencia del médico y la situación y valores del paciente (5). Este libro trata sobre varios aspectos de la medicina

booksmedicos.org

basada en la evidencia, especialmente sobre la evaluación crítica de la evidencia acerca de las preguntas clínicas.

Tabla 1-3

Factores diferentes a la medicina basada en la evidencia que pueden influir en las decisiones clínicas	
Medicina basada en la eminencia	Los compañeros más experimentados creen que la experiencia triunfa sobre las pruebas
Medicina basada en la vehemencia	Las pruebas son sustituidas por el volumen y la estridencia
Medicina basada en la elocuencia (o en la elegancia)	Elegancia en el vestir y elocuencia verbal
Medicina basada en la Providencia	Lo mejor es dejar la decisión en manos del Todopoderoso
Medicina basada en el retraimiento	Demasiado tímido para tomar cualquier decisión médica
Medicina basada en el miedo	El miedo a ser denunciado y juzgado es un potente estímulo para la sobreinvestigación y el sobretratamiento
Medicina basada en la confianza	Fanfarronadas o bravuconadas

Adaptado de Isaacs D, Fitzgerald D. Seven alternatives to evidencebased medicine. BMJ 1999;319:1618.

En los ámbitos clínicos de la vida real, existe otro tipo de «evidencia» que compite por atraer la atención de los médicos y que puede influir en las decisiones médicas. La tabla 1-3 describe algunos de ellos en una parodia sobre la medicina basada en la evidencia que fue publicada hace algunos años, pero que aún es válida hoy día. ¡Probablemente todos los médicos han experimentado al menos uno de estos factores durante sus años de apredizaje! Otro factor no tan gracioso pero muy importante se ha descrito como el nivel de evidencia IV (6). Los médicos tienden a recordar los casos donde los resultados fueron terriblemente malos en la atención que le ofrecieron a un paciente particular y es más probable que cambien su práctica tras esta

booksmedicos.org

experiencia que después de leer un estudio bien hecho. Las alternativas menos válidas que la medicina basada en la evidencia pueden ser muy atractivas a nivel emocional y pueden ser una forma conveniente de hacer frente a la incertidumbre, pero son un sustituto débil de la evidencia obtenida de una buena investigación.

La **investigación de servicios de salud** estudia la forma en que los factores no biológicos (como el personal y las instalaciones, la manera de organizar y financiar el cuidado asistencial, las creencias de los médicos y la cooperación de los pacientes) afectan a la salud de los pacientes. Estos estudios han demostrado, por ejemplo, que la asistencia médica difiere de forma sustancial de una pequeña región geográfica a otra (sin las correspondientes diferencias en cuanto a la salud de los pacientes); que la cirugía llevada a cabo en hospitales que practican un determinado procedimiento a menudo suele ofrecer mejores resultados que la cirugía en los hospitales en los que dicho procedimiento se realiza con poca frecuencia; y que un número relativamente pequeño de pacientes con cardiopatía toma ácido acetilsalicílico durante las horas posteriores al inicio del dolor torácico, a pesar de que esta sencilla práctica ha demostrado reducir el número de episodios vasculares posteriores cerca de un 25 %. Este tipo de estudios guía a los médicos en sus esfuerzos por aplicar el conocimiento existente sobre las mejores prácticas clínicas.

Existen otras ciencias que también orientan la asistencia a los pacientes. La **toma de decisiones cuantitativa** incluye **análisis de costo-efectividad**, que describen los costos económicos necesarios para lograr buenos resultados como evitar la muerte o la enfermedad, y **análisis de decisión**, que exponen la base racional para las decisiones clínicas y las consecuencias de las elecciones. Las **ciencias sociales** describen la forma en que el entorno social afecta a los comportamientos relacionados con la salud y a la utilización de los servicios sanitarios.

Las **ciencias biológicas**, que estudian la secuencia de los acontecimientos biológicos que conducen de la salud a la enfermedad, son potentes herramientas que permiten conocer cómo pueden actuar los fenómenos clínicos en los seres humanos. Históricamente, fue sobre todo el progreso de las ciencias biológicas el que determinó el enfoque científico de la medicina clínica y aún siguen desempeñando un papel fundamental. La anatomía explica los síndromes de compresión neural, su causa y su alivio. La fisiología y la bioquímica constituyen la base para el tratamiento de la

booksmedicos.org

cetoacidosis diabética. La genética molecular predice la aparición de enfermedades que van desde las cardiopatías comunes y el cáncer hasta metabolopatías congénitas raras, como la fenilcetonuria y la fibrosis quística.

Sin embargo, la comprensión de la biología de la enfermedad no suele ser, por sí misma, una base sólida para la predicción en los seres humanos sanos. Existen demasiados factores que contribuyen a la salud y a la enfermedad. En primer lugar, puede que los mecanismos de la enfermedad no se entiendan completamente. Por ejemplo, la idea de que la glucemia en los pacientes diabéticos se ve más afectada por la ingestión de azúcares simples (sacarosa o azúcar de mesa) que por azúcares complejos como el almidón (presente en las patatas y la pasta) ha sido disipada por rigurosos estudios que han comparado los respectivos efectos de ambos azúcares. Asimismo, cada vez está más claro que los efectos de las anomalías genéticas puntuales se ven a menudo modificados por el entorno genético y por un entorno físico y social complejo, en el que se incluye la alimentación de la persona, la exposición a agentes químicos e infecciosos.

Por ejemplo, la glucosa-6-fosfato deshidrogenasa (G6PD) es una enzima que protege a los eritrocitos contra las lesiones oxidativas que causan hemólisis. La deficiencia de G6PD es la deficiencia enzimática más habitual en humanos y ocurre debido a ciertas mutaciones del gen *G6PD* ligado al cromosoma X. Sin embargo, los varones con variantes genéticas comunes de la deficiencia de G6PD suelen ser asintomáticos, por lo que desarrollan hemólisis e ictericia sólo cuando se exponen al estrés oxidativo ambiental, como ciertos medicamentos o infecciones. Finalmente, como se muestra en el ejemplo del tratamiento con rosiglitazona en pacientes con diabetes tipo 2, los medicamentos pueden tener múltiples efectos sobre la salud del paciente, además de los esperados mediante el estudio de la biología de la enfermedad. Por lo tanto, el conocimiento de la biología de la enfermedad produce hipótesis, a menudo muy buenas, sobre qué podría ocurrirles a los pacientes. Sin embargo, dichas hipótesis tienen que ser demostradas mediante sólidos estudios con seres humanos sanos, antes de aceptarlas como hechos clínicos.

En resumen, la epidemiología clínica es una de las muchas ciencias básicas en la medicina clínica. A lo sumo, las diversas ciencias relacionadas con la salud se complementan entre sí. Los descubrimientos en una de ellas se confirman en la otra; los descubrimientos en la otra conducen a hipótesis nuevas en la primera.

booksmedicos.org

Ejemplo

En 1980, los médicos de San Francisco detectaron infecciones y cánceres inusuales en hombres homosexuales, enfermedades que antes se habían visto sólo en pacientes severamente inmunodeprimidos. El nuevo síndrome se denominó síndrome de inmunodeficiencia adquirida (SIDA). Los epidemiólogos establecieron que los hombres estaban sufriendo una enfermedad transmisible que afectaba tanto a hombres como a mujeres y se transmitía no sólo mediante la actividad sexual sino al compartir agujas y productos sanguíneos. Los científicos de laboratorio identificaron el virus de la inmunodeficiencia humana (VIH) y desarrollaron fármacos dirigidos específicamente a la estructura y al metabolismo de este virus. Se han desarrollado fármacos prometedores en base a la comprensión de los mecanismos biológicos y se ha probado su efectividad en ensayos clínicos. Ha surgido una nueva especialidad clínica para la atención de los pacientes con infección por VIH. Los trabajadores de salud pública han promovido el sexo seguro y otros programas para prevenir la infección por VIH. Por lo tanto, los médicos, epidemiólogos, científicos de laboratorio y trabajadores de salud pública han contribuido al control de esta nueva enfermedad, sobre todo en los países más desarrollados, causando un aumento importante en la supervivencia y en mejorar la calidad de vida de las personas infectadas por el VIH.

PRINCIPIOS BÁSICOS

El objetivo de la epidemiología clínica es contribuir a la creación de métodos de observación e interpretación clínicas que den lugar a conclusiones válidas y conduzcan a proporcionar una mejor asistencia al paciente. Las respuestas más creíbles a las preguntas clínicas se basan en un reducido número de principios básicos. Dos de ellas, que las observaciones deben abordar las preguntas que afrontan los pacientes y los médicos, y que los resultados deben incluir los resultados sobre la salud centrados en el paciente (las 5 D), ya han sido cubiertas. Otros principios básicos se discuten a continuación.

Variables

Los investigadores denominan **variables** (aquello que varía y puede medirse) a los atributos de los pacientes y a los episodios clínicos. En un estudio típico, existen dos tipos de variables. Una es una supuesta causa o variable predictora, a veces denominada **variable independiente**; la otra es un posible efecto, llamada en ocasiones **variable dependiente**. Existen otras variables que pueden formar parte del sistema estudiado y que pueden afectar a la relación entre estas dos. Dichas variables se denominan **variables extrañas** (o **covariables**) porque son ajenas a la pregunta principal (aunque puede que estén muy relacionadas con el fenómeno estudiado).

booksmedicos.org

Números y probabilidad

La ciencia clínica, como todas las ciencias, depende de determinaciones cuantitativas. Las impresiones personales, el instinto y las creencias también son importantes en medicina, pero sólo como complementos de una base sólida de información numérica. Esta base permite una mejor confirmación, una comunicación más precisa entre médicos y entre médicos y pacientes, y una estimación de los errores. Los resultados clínicos, como la aparición de enfermedades, la muerte, los síntomas o la incapacidad, pueden expresarse en forma de números.

En la mayoría de las situaciones clínicas, el diagnóstico, el pronóstico y los resultados del tratamiento son inciertos para un paciente concreto. Un paciente podrá presentar un resultado clínico o no; la predicción de este resultado rara vez es exacta. Por lo tanto, la predicción debe expresarse como una probabilidad. Las probabilidades para un paciente concreto se calculan mejor si se tiene en cuenta la experiencia con grupos de pacientes similares: por ejemplo, que el hábito de fumar cigarrillos duplica el riesgo de morir a cualquier edad, que los análisis de sangre en los que se miden las troponinas detectan cerca del 99 % de los infartos de miocardio en pacientes con dolor torácico agudo, y que entre el 2 % y el 6 % de los pacientes sometidos a una intervención quirúrgica por aneurisma aórtico abdominal morirá en los 30 días posteriores a la intervención electiva o programada en comparación con el 40-80 % cuando es necesaria una reparación urgente.

Poblaciones y muestras

Las **poblaciones** son todas las personas que pertenecen a un ámbito definido (como Carolina del Norte) o que reúnen unas características definidas (como ser mayores de 65 años o presentar un nódulo tiroideo). La población habitual de los estudios epidemiológicos suele estar compuesta por personas no seleccionadas de la comunidad. Por otro lado, las poblaciones clínicas están formadas por todos los pacientes que presentan una característica clínica concreta, como una neumonía extrahospitalaria o una estenosis aórtica. Así pues, se habla de población general, población hospitalizada o población de pacientes con una enfermedad concreta.

Una **muestra** es un subgrupo de personas en una población definida. Habitualmente, la investigación clínica se lleva a cabo en muestras. Puede que nos interesen las características de una población definida, pero, por

booksmedicos.org

razones prácticas, sólo podemos (y debemos) deducirlas mediante la descripción de las características que presentan las personas de una muestra (fig. 1-2). A partir de este punto, realizamos la **inferencia**, un juicio razonado basado en datos, de que las características de la muestra representan a las características de la población.

La medida en que una muestra representa a su población, y por lo tanto la sustituye de forma correcta, depende del modo en que se seleccionó la muestra. Los métodos por los que cada miembro de la población tiene posibilidades equivalentes (o conocidas) de ser seleccionado pueden dar lugar a muestras extraordinariamente similares a la población de origen, al menos a largo plazo y para muestras de gran tamaño. Un ejemplo cotidiano de ello son los sondeos de opinión que realizan muestreos de los hogares a partir de los datos censales. En la investigación clínica que nosotros llevamos a cabo, a menudo nos servimos de una computadora para seleccionar una muestra representativa de todos los pacientes que acuden a nuestra gran consulta multidisciplinaria compartida, cada uno de los cuales tiene las mismas posibilidades de ser seleccionado. Por otro lado, las muestras tomadas de forma descuidada o por propia conveniencia (es decir, seleccionando pacientes con los que resulta fácil trabajar, o porque da la casualidad de que acuden a la consulta los días en los que se recogen los datos) pueden representar de forma errónea a la población de origen y, por lo tanto, dar lugar a resultados erróneos.

Figura 1-2 ■ Población y muestra.

Sesgo (error sistemático)

booksmedicos.org

El **sesgo** es «un proceso que, en cualquier etapa de una inferencia, tiende a producir resultados que se desvían sistemáticamente de los valores verdaderos» (7). Se trata de «cualquier tendencia en la recolección, análisis, interpretación, publicación o revisión de los datos que puede dar lugar a conclusiones que difieren sistemáticamente (en lugar de ser aleatoria) de la verdad» (8).

Ejemplo

Los pacientes con hernia inguinal que se someten a una intervención laparoscópica presentan aparentemente menos dolor postoperatorio y se reincorporan antes a la vida laboral que los pacientes sometidos a la cirugía abierta tradicional. El médico prudente se preguntará: «Los resultados de la cirugía laparoscópica, ¿son realmente mejores o sólo lo parecen por la presencia de sesgos en el modo en que se recogió la información?» Puede que la intervención laparoscópica sólo se ofrezca a los pacientes que gozan de mejor salud, o que parecen tener una resistencia tisular mayor, debido a su edad o su estado de salud general. Quizá los cirujanos y los pacientes se sienten más inclinados a pensar que el procedimiento causará menos dolor, porque se trata de una técnica nueva y el tamaño de la cicatriz será menor, de forma que los pacientes señalan un grado de dolor menor y existen menos probabilidades de que los cirujanos pregunten por el mismo o que lo registren en la historia clínica. Puede que a los pacientes que se someten a cirugía laparoscópica se les indique que volverán al trabajo antes que a los que se someten a cirugía abierta. En caso de que se hubiera dado alguna de estas circunstancias, los resultados favorables podrían estar relacionados con diferencias sistemáticas en el modo de seleccionar a los pacientes para el procedimiento laparoscópico, en la forma en que estos notificaron sus síntomas o en la manera en que se les indicó lo que podían hacer, en lugar de radicar en una verdadera diferencia en el número de éxitos. Como se discute en el capítulo 5, hay formas de protegerse contra los posibles sesgos. Los estudios que han evitado estos sesgos han detectado que los pacientes que se someten a una cirugía laparoscópica sí presentan, de hecho, menos dolor y regresan antes al trabajo en pocos días. Pero la cirugía laparoscópica tarda más tiempo y varios estudios encontraron complicaciones más graves en los pacientes a los que se les realizó esta intervención, así como una mayor tasa de recurrencia, sobre todo entre los hombres mayores (9,10). En resumen, los estudios cuidadosos encontraron que la elección entre ambos procedimientos no está clara.

Tabla 1-4

Sesgos en las observaciones clínicas	
Sesgo de selección	Se produce cuando se comparan grupos de pacientes que difieren en factores determinantes para el resultado que no son objeto del estudio
Sesgo de medición	Se produce cuando los métodos de medición difieren entre los grupos de pacientes

booksmedicos.org

Sesgo de confusión	Se produce cuando dos factores están asociados (discurren juntos) y el efecto de uno se confunde o queda distorsionado por el efecto del otro

Las observaciones de los pacientes (ya sea para su asistencia o para la investigación) son especialmente propensas a sufrir sesgos. El proceso tiende a ser sencillamente caótico. Como participantes en un estudio, los seres humanos tienen el desconcertante hábito de hacer lo que les place, algo que no coincide necesariamente con lo que deberían hacer para producir respuestas rigurosamente científicas. Cuando los investigadores intentan realizar algún experimento con ellos, como lo harían en un laboratorio, las cosas tienden a salir mal. Algunas personas rehúsan participar, mientras que otras abandonan o eligen otro tratamiento. Además, los médicos se sienten inclinados a creer que sus tratamientos son satisfactorios. (La mayoría de los pacientes no estarían dispuestos a que les atendiera un médico que pensara lo contrario.) Esta actitud, tan importante en la práctica de la medicina, hace que las observaciones clínicas sean especialmente vulnerables a los sesgos.

Aunque se han definido docenas de sesgos (11), la mayor parte se puede incluir en una de tres grandes categorías (tabla 1-4).

Sesgo de selección

El **sesgo de selección** se produce cuando se comparan grupos de pacientes que, además de diferir en los factores principales del estudio, también difieren en otros que pueden afectar al resultado del mismo. Los grupos de pacientes difieren, a menudo, en muchos aspectos: edad, sexo, gravedad de la enfermedad, presencia de otras enfermedades, asistencia que reciben, etc. Si se comparan dos grupos que difieren en una característica específica que resulta de interés (p. ej., un tratamiento o una posible causa de enfermedad), pero que además presentan diferencias en esos otros aspectos comentados, y dichas diferencias están relacionadas directamente con el resultado, la comparación estará sesgada y poca cosa se podrá concluir de los efectos independientes de la característica de interés. En el ejemplo de la herniorrafia, se habría producido el sesgo de selección si los pacientes sometidos al procedimiento laparoscópico hubieran presentado un mejor estado de salud que los pacientes sometidos a cirugía abierta.

Sesgo de medición

booksmedicos.org

El **sesgo de medición** se produce cuando los métodos de medición dan lugar a resultados sistemáticamente incorrectos.

Ejemplo

Los niveles de presión sanguínea son un predictor poderoso de enfermedad cardiovascular. Sin embargo, múltiples estudios han demostrado que medir la presión sanguínea no es tan simple como parece (12). La medición correcta requiere del uso de procedimientos apropiados, incluyendo el uso de un manguito de tamaño mayor en adultos con sobrepeso y obesidad, cuidar la posición del paciente de tal manera que la porción superior del brazo esté por debajo del nivel de la aurícula derecha y que el paciente no tenga que sostener su brazo, tomar la medida en un ambiente tranquilo, haciéndolo varias veces. Si cualquiera de estos procedimientos no se realiza de forma correcta, es probable que las mediciones resultantes estén artificial y sistemáticamente elevadas. Otro factor que puede provocar lecturas de la presión arterial sistemáticamente más altas, en ocasiones denominada «hipertensión de bata blanca» (fig. 1-3), ocurre cuando los médicos miden la presión arterial, sugiriendo que las visitas al médico provocan ansiedad en los pacientes. No obstante, es probable que los médicos que desinflan el manguito para medir la presión arterial más rápido de 2-3 mm/s subestimarán la presión sistólica (pero sobreestimarán la diastólica). Los estudios han mostrado que la tendencia de los médicos es registrar valores que están justo en el nivel normal en pacientes con presión arterial en el límite alto. Por lo tanto, los errores sistemáticos en la medición de la presión arterial pueden conducir al tratamiento excesivo o insuficiente de los pacientes en la práctica clínica. La investigación clínica basada en las mediciones de la presión arterial tomada durante la atención rutinaria de los pacientes puede dar lugar a resultados erróneos a menos que se utilicen procedimientos estandarizados cuidadosos. Este tipo de sesgos condujeron al desarrollo de instrumentos de medición de la presión arterial que no requieren del oído o de las manos humanas.

Confusión

La **confusión** puede ocurrir cuando el investigador intenta averiguar si un factor, como por ejemplo una conducta o la exposición a un fármaco, es causa de enfermedad por sí mismo o si está relacionado con esta. Si ese factor se asocia o «viaja junto» a otro factor, que a su vez está relacionado con el resultado, el efecto del primero puede confundirse con el del segundo o quedar distorsionado por este.

booksmedicos.org

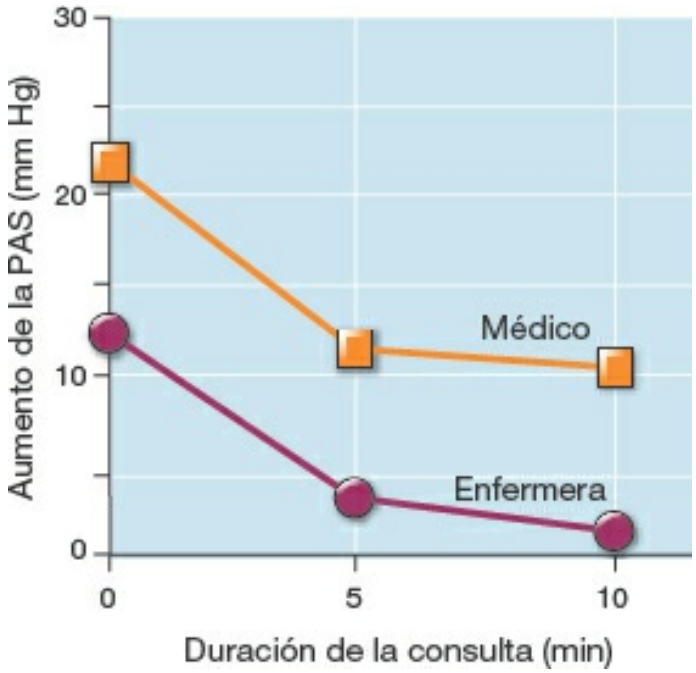


Figura 1-3 ■ **Hipertensión de bata blanca.** Aumento en la presión arterial sistólica (PAS) determinada por la monitorización continua intraarterial, si la presión sanguínea se mide con un esfigmomanómetro por un médico o enfermera desconocidos. (Redibujado con autorización de Mancia G, Parati G, Pomidossi G, et al. Alerting reaction and rise in blood pressure during measurement by physician and nurse. Hypertension 1987;9:209-215.)

Ejemplo

Los suplementos de antioxidantes, como las vitaminas A, C y E, son populares entre el público en general. Los experimentos de laboratorio y los estudios de personas que eligen tomar antioxidantes sugirieron que los antioxidantes previenen la enfermedad cardiovascular y ciertos tipos de cáncer. Sin embargo, los estudios aleatorizados minuciosos, que son capaces de evitar el sesgo de confusión, habitualmente encontraron poco efecto de los antioxidantes (13,14). De hecho, cuando se combinaron los resultados de estos estudios, el uso de antioxidantes, especialmente en dosis elevadas, se asoció con un pequeño aumento, y no con la disminución, de las tasas de mortalidad. ¿Cómo podrían los resultados de los primeros estudios armonizar con los resultados opuestos de los recientes ensayos cuidadosamente controlados? Se ha sugerido el sesgo de confusión, como se ilustra en la figura 1-4. Es probable que las personas que toman antioxidantes por iniciativa propia hagan cosas diferentes que quienes no toman antioxidantes, como ejercitarse más, controlar su peso, comer más vegetales y no fumar, y pueden ser estas actividades, y no los antioxidantes, que disminuyen las tasas de muerte en los estudios donde no hubo aleatorización de la intervención.

booksmedicos.org

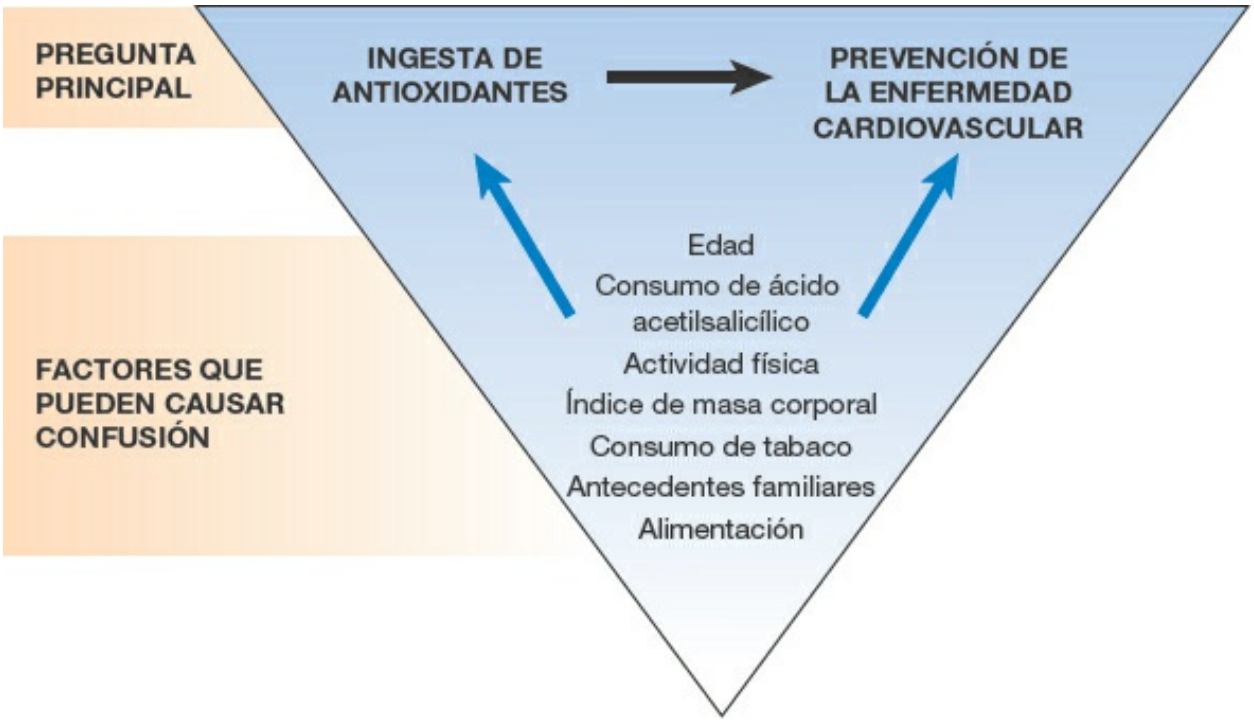


Figura 1-4 ■ Confusión. La relación entre el consumo de antioxidantes y el riesgo cardiovascular es un factor de confusión debido a las características y la conducta del paciente relacionados con el uso de antioxidantes y el desarrollo de enfermedad cardiovascular.

La mayoría de los estudios de investigación clínica, especialmente los que observan a las personas durante un tiempo, a menudo intentan evitar la confusión al «controlar» las posibles variables de confusión durante el análisis (v. cap. 5). Las variables como la edad, el sexo y la raza casi siempre son analizadas como confusores porque muchos resultados sobre la salud varían de acuerdo con estos. Los estudios que afectan a la conducta humana (como tomar antioxidantes regularmente) son especialmente susceptibles al sesgo de confusión porque esta conducta es tan compleja que es difícil analizar todos los factores que pueden influir sobre ella.

Una variable no es confusora si se halla directamente en el trayecto que conduce de la causa al efecto. Una variable de confusión tampoco tiene por qué ser una causa por sí misma. Puede que esté relacionada con la posible causa y con el efecto en un conjunto concreto de datos que estén a mano, debido a un sesgo de selección o al azar, pero no tiene por qué estar relacionada con los primeros por naturaleza. Tanto si la relación se encuentra en los datos como en su naturaleza, las consecuencias son las mismas: la impresión errónea de que el factor de interés constituye una causa independiente y real cuando en realidad no lo es.

booksmedicos.org

Los sesgos de selección y de confusión no son mutuamente excluyentes; sin embargo, se describen por separado porque presentan problemas en distintos puntos de un estudio clínico. El sesgo de selección debe tenerse en cuenta cuando se seleccionan pacientes para una investigación y cuando se realiza el diseño de un estudio. El sesgo de confusión debe abordarse durante el análisis de los datos, una vez que se han realizado las observaciones. Se produce más de un sesgo en el mismo estudio, como en el siguiente ejemplo hipotético.

Ejemplo

Ha surgido la preocupación de que el consumo de cafeína durante el embarazo puede producir resultados fetales adversos. No sería ético determinar si la cafeína es peligrosa para los fetos en un experimento en el que se diera a algunas mujeres embarazadas niveles elevados de cafeína, mientras que a otras no, por lo que los investigadores generalmente han estudiado lo que ocurre durante el embarazo según la cantidad de cafeína ingerida. Sin embargo, se han demostrado varios sesgos en muchos de estos estudios (15). Es posible que existieran sesgos de medición porque la mayoría de los estudios se basaron en la ingesta de cafeína que los pacientes reportaron. Un estudio mostró el sesgo de recuerdo, un tipo de sesgo de medición que se refiere a las diferencias del recuerdo entre las personas con un resultado adverso cuando se comparan con aquellos con un resultado normal. Se encontró una asociación entre el consumo de cafeína y el aborto cuando se entrevistó a las mujeres después de haber abortado, pero no cuando fueron interrogadas sobre el consumo de cafeína antes del aborto (16). Si se reclutaran algunas mujeres para estudios de cafeína durante las consultas prenatales (mujeres que tienen mayor probabilidad de ser particularmente conscientes sobre su salud) y otras se incorporan hacia el final del embarazo, las diferentes estrategias de reclutamiento podrían conducir a un sesgo de selección que podrían invalidar los resultados. Finalmente, se sabe que el consumo excesivo de café se asocia con el tabaquismo, niveles socioeconómicos más bajos, mayor consumo de alcohol y en general con menos conciencia sobre la salud; todo esto puede confundir cualquier asociación entre la cafeína y los resultados fetales adversos.

El hecho de que se puedan producir sesgos no implica que en un estudio concreto realmente los haya o que, si los hubiera, ejerzan una gran influencia sobre los resultados que interesan. Para que un investigador o un lector pueda abordar un sesgo de forma eficaz, primero necesita saber dónde y cómo buscarlo, y si lo encuentra, debe saber qué hacer con él. También es necesario determinar si realmente se ha producido un sesgo y cuáles son sus posibles dimensiones, tras lo cual hay que decidir si es lo bastante grave como para modificar las conclusiones del estudio de una forma clínicamente importante.

booksmedicos.org

Azar

A menudo, las observaciones sobre una enfermedad se realizan en una muestra de pacientes, porque resulta imposible estudiar a todos los pacientes con dicha enfermedad. Estas muestras, incluso si se ha evitado el sesgo en la selección, pueden representar de forma errónea la situación en la población como un todo a causa del **azar**. A pesar de todo, los resultados de una muestra sin sesgo tienden a aproximarse al valor real. Si se repitieran las observaciones en muchas muestras de pacientes de este tipo, los resultados se concentrarían alrededor del valor real, con la mayoría de dichos valores más bien próximos que alejados del valor real. La divergencia entre la observación de una muestra y el valor real de la población, causada únicamente por el azar, se denomina **variación aleatoria**.

Todos conocemos el azar como la razón que explica que, si se lanza una moneda al aire, supongamos, 100 veces, no salga cara el 50 % de ellas. El mismo efecto, la variación aleatoria, se aplica cuando se comparan los efectos de las reparaciones abierta y laparoscópica de la hernia inguinal. Supongamos que se eliminan todos los sesgos de un estudio sobre los efectos de ambos procedimientos y que estos son igual de eficaces, y cada uno de ellos va seguido de dolor en el 10 % de los pacientes. Aun así, debido únicamente al azar, un solo estudio con un número reducido de pacientes en cada grupo de tratamiento podría hallar fácilmente que los pacientes se recuperan de forma más favorable con la laparoscopia que con la cirugía abierta (o viceversa).

El azar puede afectar a todas las etapas de una observación clínica. En la evaluación de los dos métodos de reparación de la hernia inguinal, la variación aleatoria se produce en el muestreo de los pacientes para el estudio, en la selección de los grupos de tratamiento, y en las mediciones del dolor y del retorno al trabajo.

Al contrario que el sesgo, que tiende a distorsionar la situación en una dirección o en otra, es tan probable que la variación aleatoria dé lugar a observaciones situadas por encima como por debajo del valor real. Como consecuencia, la media de numerosas observaciones de muestras realizadas sin la presencia de sesgos tiende a aproximarse al valor real de la población, aunque pueda no suceder lo mismo con los resultados de pequeñas muestras concretas. En el caso de la reparación de la hernia inguinal, la evaluación de múltiples estudios en conjunto ha mostrado que los resultados de la reparación laparoscópica son menos dolorosos en los primeros días después

booksmedicos.org

de la cirugía.

La estadística puede ser útil para calcular hasta qué punto el azar (variación aleatoria) es responsable de los resultados de un estudio clínico. Los conocimientos de estadística también pueden contribuir a reducir el papel que desempeña el azar, ya que permiten crear mejores diseños y análisis. Sin embargo, la variación aleatoria nunca puede eliminarse del todo, por lo que durante la valoración de los resultados de las observaciones clínicas siempre debe tenerse en cuenta el azar, cuyo papel en las observaciones clínicas se comenta en el capítulo 11.

Los efectos del sesgo y del azar son acumulativos

Las dos fuentes de errores (el sesgo y el azar) no son mutuamente excluyentes; en la mayor parte de las situaciones, ambas están presentes. La relación entre ellas se ilustra en la figura 1-5. Como ejemplo, se ha considerado la medición de la presión arterial diastólica en un solo paciente; cada punto representa una observación de dicho enfermo. La presión arterial real de este paciente, 80 mm Hg, se obtiene mediante una cánula intraarterial, un método que no es factible para las mediciones sistemáticas. Habitualmente, se mide de forma indirecta, con un esfigmomanómetro (o tensiómetro, aparato provisto de un manguito). Como se discutió en un ejemplo anterior, este instrumento más sencillo es propenso al error o a las desviaciones respecto al valor real. En la figura, el error está representado por todas las lecturas del esfigmomanómetro situadas a la derecha del valor real. La desviación de estas lecturas a valores altos (sesgo) puede deberse a varios motivos (p. ej., un tamaño de manguito erróneo, un paciente con ansiedad o un médico sordo). Las lecturas individuales de la presión arterial también están sujetas a error debido a la variación aleatoria en las mediciones, como ilustra la dispersión de las lecturas del esfigmomanómetro alrededor del valor medio (90 mm Hg).

booksmedicos.org

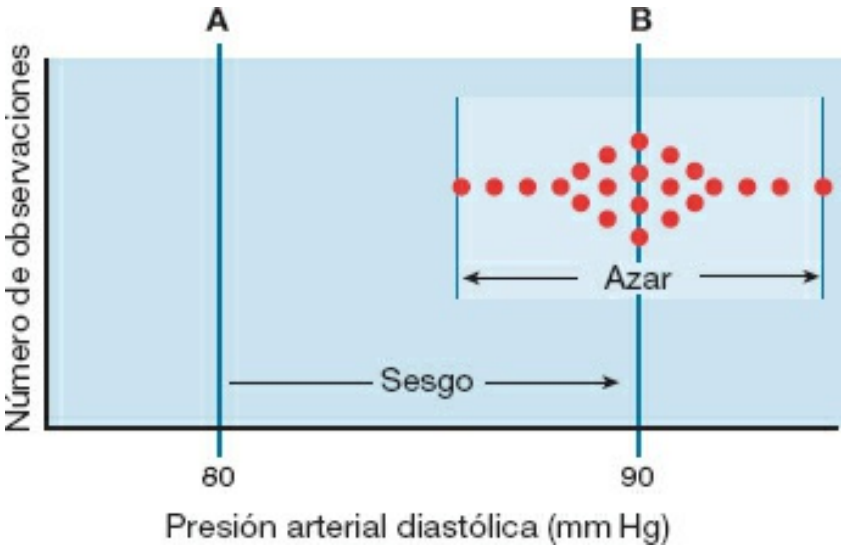


Figura 1-5 ■ Sesgo y azar. Presión arterial real medida con una cánula intraarterial **(A)** y medición clínica realizada con un esfigmomanómetro **(B).**

La razón principal para distinguir entre sesgo y azar es que estos deben manejarse de forma diferente. En teoría, el sesgo puede evitarse si las investigaciones clínicas se llevan a cabo adecuadamente o puede corregirse mediante el análisis idóneo de los datos. Si no se elimina, el sesgo es detectado a menudo por el lector crítico. La mayor parte de este libro gira en torno al modo de reconocer, evitar o minimizar el sesgo. Por otro lado, el azar no puede eliminarse, si bien su influencia puede reducirse mediante un diseño adecuado de la investigación, y su efecto, calcularse mediante la estadística. Ningún tratamiento estadístico puede corregir los sesgos de origen desconocido en los datos. Algunos estadísticos hasta sugieren que la estadística no debería aplicarse a los datos que fueran vulnerables al sesgo a causa de un mal diseño de la investigación, ya que podría otorgarse un falso respeto a un trabajo que fundamentalmente induce a error.

Validez interna y externa

Cuando realizan inferencias sobre una población a partir de las observaciones de una muestra, los médicos deben tener clara la respuesta a dos preguntas básicas: en primer lugar, ¿son correctas las conclusiones de la investigación para las personas de la muestra?, y, segundo, en caso afirmativo, ¿representa la muestra adecuadamente a los pacientes que revisten más interés, como el tipo de pacientes que suelen atender, o quizá un paciente concreto que tengan a mano? (fig. 1-6).

booksmedicos.org

La **validez interna** es el grado en que los resultados de un estudio son correctos para la muestra de pacientes estudiados. Es «interna» porque se aplica a las condiciones del grupo concreto de pacientes que se está observando y no necesariamente a otros. La validez interna de una investigación clínica viene determinada por lo bien que se realizaron el diseño del estudio, la recogida de datos y los análisis, y se ve amenazada por todos los sesgos y la variación aleatoria descritos anteriormente. Para que una observación clínica resulte útil, la validez interna es necesaria, aunque no es una condición suficiente.

La **validez externa** es el grado en que los resultados de una observación son válidos en otros ámbitos. Otro término para el mismo concepto es capacidad de **generalización**. Para un médico concreto, se trata de la respuesta a la siguiente pregunta: «Suponiendo que los resultados de un estudio son reales, ¿son también aplicables a mis pacientes?» La capacidad de generalización expresa la validez de suponer que los pacientes de un estudio son parecidos a otros pacientes.

Cada estudio es generalizable al resto de los pacientes en la misma medida en que lo es a los pacientes del estudio. Sin embargo, un estudio impecable, con una elevada validez interna, puede inducir por completo a error si sus resultados se generalizan en los pacientes equivocados.

booksmedicos.org

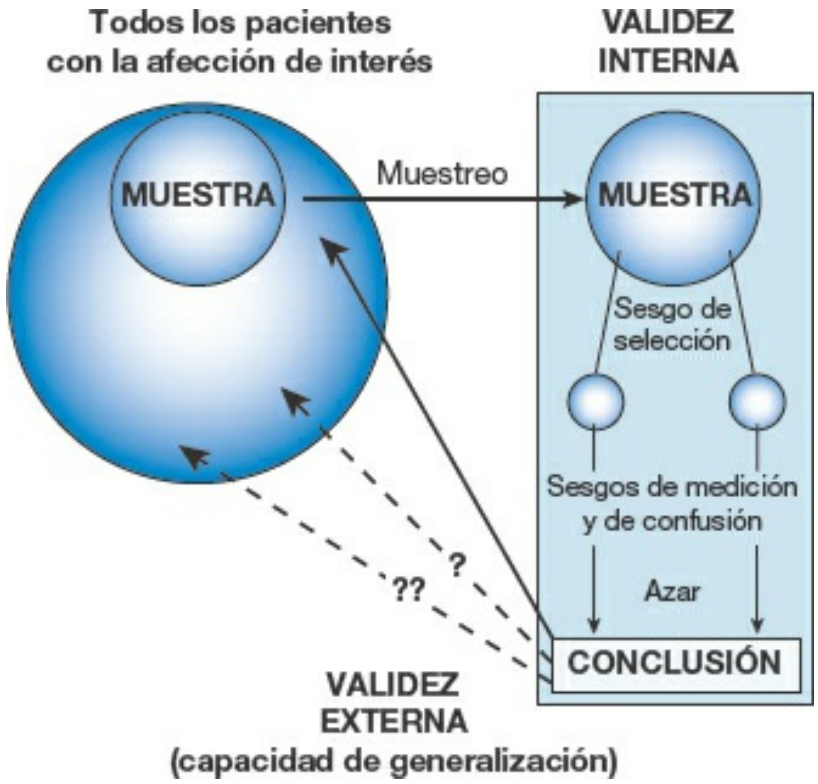


Figura 1-6 ■ Validez interna y externa.

La capacidad de generalización de las observaciones clínicas, incluso de aquellas con una elevada validez interna, es una cuestión de juicio personal de la que incluso las personas razonables podrían estar en desacuerdo. A menudo se produce una situación cuando los médicos deben decidir si han de utilizar los resultados de un estudio bien hecho en un paciente de más edad que los del estudio, de diferente género o más enfermo. Es probable que un tratamiento que funciona bien en hombres jóvenes adultos cause más daño que bienestar en mujeres mayores y más enfermas.

Ejemplo

En la anorexia nerviosa, un trastorno de la alimentación que afecta fundamentalmente a mujeres jóvenes, ¿cuál es el índice de mortalidad a largo plazo? En una síntesis de 42 estudios, la mortalidad estimada fue del 15 % en 30 años (17). Estos estudios, como la mayoría de las investigaciones clínicas, se realizaron con pacientes identificadas en centros de referencia, donde se tratan casos relativamente graves. Un estudio de todas las pacientes que sufren anorexia en una población definida arrojó una perspectiva diferente de la enfermedad. Los investigadores de la Clínica Mayo pudieron identificar a todas las pacientes de su ciudad, Rochester, Minnesota (EE.UU.), que sufrieron esta enfermedad entre 1935 y 1989 (fig. 1-7) (18). La mortalidad por

booksmedicos.org

todas las causas a los 30 años fue del 7 %, la mitad de la notificada en los estudios. La mortalidad prevista para estas pacientes, teniendo en cuenta su edad y sexo, fue casi la misma, el 6 %. Por lo tanto, aunque algunas pacientes sí mueren a causa de la anorexia nerviosa, los estudios publicados sobreestiman el riesgo en gran medida, presumiblemente porque notifican su experiencia con casos relativamente graves.

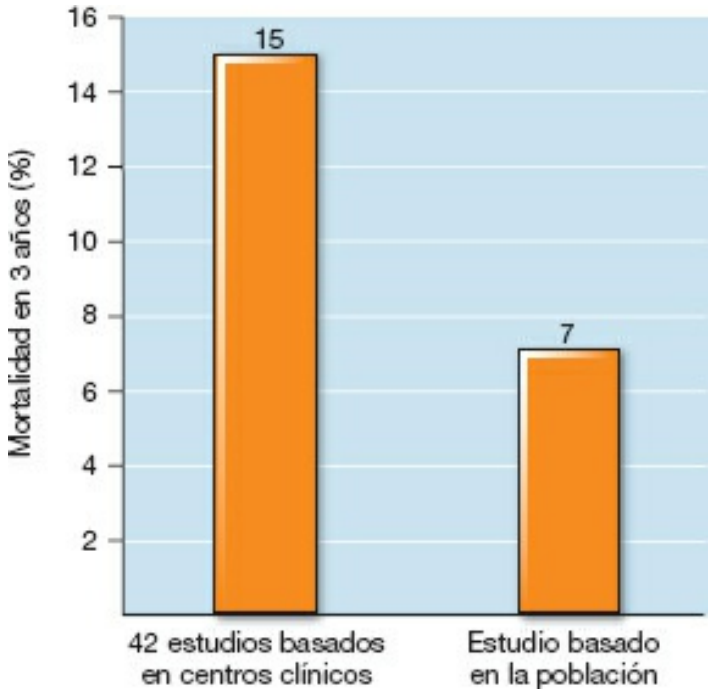


Figura 1-7 ■ Sesgo de muestreo. Mortalidad por todas las causas durante 30 años en pacientes con anorexia nerviosa. Comparación entre una síntesis de 42 estudios publicados, la mayoría realizados en centros clínicos, y un estudio de todas las pacientes con anorexia nerviosa realizado en la población. (Datos de Sullivan PF. Mortality in anorexia nervosa. Am J Psychiatry 1995;152:1073–1074; and Korndorter SR, Lucan AR, Suman VJ, et al. Long-term survival of patients with anorexia nervosa: a population-based study in Rochester, Minn. Mayo Clin Proc 2003;78:278–284.)

La capacidad de generalización muy pocas veces se puede abordar de forma satisfactoria en un solo estudio. Incluso una población definida en términos geográficos es una muestra sesgada de poblaciones de mayor tamaño. Por ejemplo, los pacientes de un hospital son muestras sesgadas de los residentes en las provincias; las provincias, de las regiones; las regiones, del estado, y así sucesivamente. Lo mejor que puede hacer un investigador con respecto a la capacidad de generalización es garantizar la validez interna, asegurarse de que la población del estudio es adecuada para el tema que se investigará, describir minuciosamente a los pacientes y evitar la inclusión de pacientes tan poco frecuentes que la experiencia obtenida con ellos sólo sea generalizable a un número reducido de otros pacientes. A partir de ahí,

booksmedicos.org

corresponde a otros estudios, en otros ámbitos, ampliar la capacidad de generalización.

INFORMACIÓN Y DECISIONES

Uno de los objetivos principales de este libro es incidir en la calidad de la información clínica y en su correcta interpretación. La toma de decisiones es otro tema. Las buenas decisiones dependen no sólo de la buena información, sino también de otros factores, como los juicios de valor y la evaluación de los riesgos y los beneficios derivados.

En estos últimos años, la toma de decisiones médicas se ha convertido en una valorada disciplina por derecho propio. Este campo comprende estudios cualitativos sobre la forma en que los médicos toman decisiones y la manera en que este proceso puede quedar sesgado y puede mejorarse. También incluye métodos cuantitativos, como el análisis de decisiones, el análisis de costo-beneficio y el análisis de costo-efectividad, que presentan el proceso de toma de decisiones de forma explícita, de modo que pueden examinarse sus componentes y las consecuencias de asignarles distintos valores y probabilidades.

Las decisiones clínicas las toman los médicos y los pacientes. En el mejor de los casos, las toman juntos en un proceso denominado **toma de decisiones compartida**, lo que revela que la pericia de ambos es complementaria. Los pacientes son expertos en aquello que esperan lograr a través de su asistencia, dadas sus preferencias y experiencias únicas. Puede que hayan encontrado abundante información sobre su enfermedad (p. ej., en Internet), pero no tienen la suficiente base para distinguir las afirmaciones creíbles de las engañosas. Los médicos son expertos en valorar si es posible alcanzar los objetivos de los pacientes y en determinar la mejor forma de lograrlo. Por este motivo, dependen del conjunto de pruebas de investigación y de la capacidad, basada en los principios de la epidemiología clínica, de distinguir entre las pruebas argumentadas y las poco convincentes. Obviamente, los médicos también aportan a la entrevista su experiencia acerca de cómo se presenta la enfermedad y las consecuencias que tendrá el tratamiento para el paciente, por ejemplo, que habrá que intubar o amputar, aspectos con los que los pacientes tienen poca experiencia. Para que los médicos puedan desempeñar su papel en este equipo, tienen que ser expertos en la interpretación de la información clínicamente pertinente.

booksmedicos.org

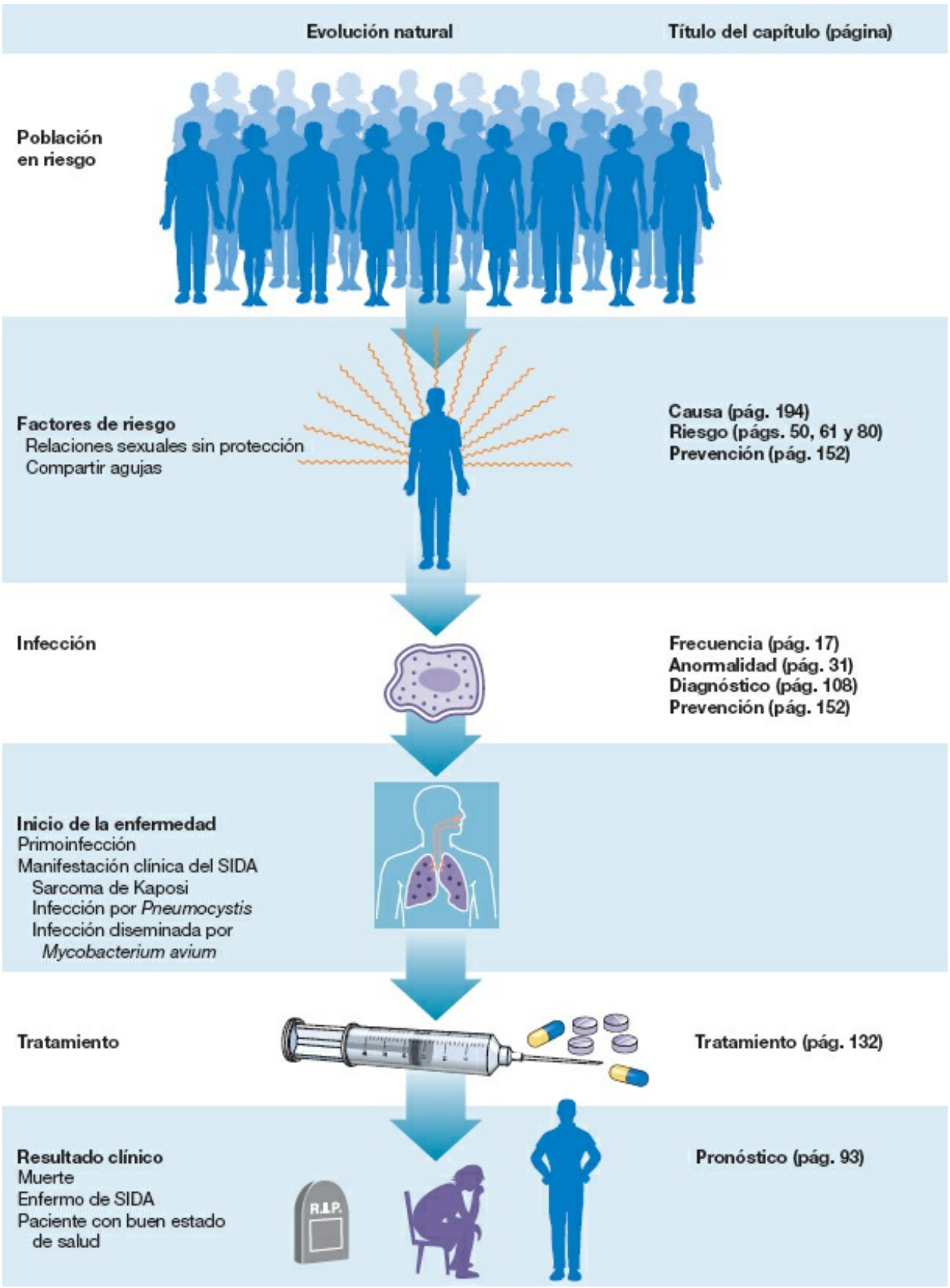
Las preferencias de los pacientes y las pruebas sólidas constituyen la base para elegir entre las opciones de tratamiento. Por ejemplo, un paciente con una valvulopatía cardíaca puede preferir la posibilidad de gozar de una buena salud a largo plazo que le ofrece la cirugía, a pesar de que esta se asocie a malestar y al riesgo de un fallecimiento a corto plazo. El médico, provisto de la habilidad de la lectura crítica y la comunicación, puede contribuir a que el paciente entienda la trascendencia de los posibles riesgos y beneficios, así como el grado de seguridad con que estos han sido establecidos.

En este libro, se incluyen algunos aspectos del análisis de decisiones, como la evaluación de las pruebas diagnósticas; sin embargo, hemos decidido no profundizar en la propia toma de decisiones médicas. La razón para ello es que las decisiones son tan buenas como la información que se ha utilizado para tomarlas, y teníamos tanto que decir sobre los aspectos clave de la recopilación y la interpretación de la información clínica como para editar, literalmente, un libro.

ORGANIZACIÓN DE ESTE LIBRO

En la mayoría de los libros de texto sobre medicina clínica, la información acerca de cada enfermedad se presenta en forma de respuestas a las preguntas clínicas tradicionales: diagnóstico, evolución clínica, tratamiento, etc. Por otro lado, la mayor parte de los libros sobre epidemiología se organizan a partir de las estrategias de investigación, como los ensayos clínicos, las encuestas, los estudios de casos y controles, etc. Esta forma de organizar un libro puede ser útil para los que realizan investigaciones clínicas, pero a menudo resulta dificultosa para los médicos.

booksmedicos.org



booksmedicos.org

Figura 1-8 ■ Organización de este libro en relación con la evolución natural de la infección por el VIH. Los capítulos 11, 13 y 14 describen temas interdisciplinares relacionados con todos los puntos de la evolución natural de la enfermedad

Para organizar esta obra, nos hemos basado principalmente en las preguntas que deben afrontar los médicos cuando atienden a sus pacientes (v. tabla 1-1). La figura 1-8 ilustra la correspondencia entre dichas preguntas y los capítulos de este libro, tomando como ejemplo la infección por el virus de la inmunodeficiencia humana (VIH). Las preguntas se relacionan con toda la evolución natural de la enfermedad, desde el momento en que las personas sin infección por el VIH se exponen por primera vez al riesgo, pasando por el período en que algunos contraen la enfermedad y se convierten en pacientes, hasta las complicaciones de esta, las enfermedades que definen al SIDA, de la supervivencia a la muerte.

En cada capítulo, hemos descrito las estrategias de investigación que se utilizan para responder a las preguntas clínicas del mismo. Algunas de estas estrategias, como los estudios de cohortes, resultan útiles para responder a varios tipos de preguntas clínicas. Para facilitar su presentación, hemos tratado cada estrategia principalmente en un capítulo, y únicamente nos referimos a ella cuando el método resulta pertinente para las preguntas de otros capítulos.

Preguntas de revisión

Las preguntas 1-6 están formuladas según el siguiente supuesto clínico.

Una mujer de 37 años de edad con lumbalgia durante las últimas 4 semanas quiere saber si usted recomendaría una intervención quirúrgica. Si es posible, usted prefiere que el tratamiento que vaya a recomendar se base en pruebas de investigación. En el estudio más valido que encuentra, los investigadores revisaron consecutivamente las historias clínicas de 40 varones con lumbalgia que eran atendidos en su consulta. Se había asignado un tratamiento quirúrgico a 22 de los pacientes, y los 18 restantes permanecieron con asistencia médica sin cirugía. El estudio comparó los índices de dolor incapacitante al cabo de 2 meses. Todos los pacientes tratados quirúrgicamente y 10 de los pacientes tratados con fármacos seguían siendo

booksmedicos.org

atendidos en la consulta al cabo de ese tiempo. Los índices de alivio del dolor fueron ligeramente superiores en los pacientes tratados quirúrgicamente.

Los siguientes enunciados representan una amenaza para la validez. Señale con qué tipo de amenaza se corresponden.

1-1. Debido a que este estudio incluyó un número relativamente reducido de pacientes, la impresión que da sobre la eficacia real de la cirugía puede ser errónea.

A. Sesgo de selección.
B. Sesgo de medición.
C. Sesgo de confusión.
D. Azar.
E. Validez externa (capacidad de generalización).

1-2. Puede que los resultados de este estudio no puedan aplicarse a su paciente, una mujer, ya que todos los pacientes del estudio eran varones.

A. Sesgo de selección.
B. Sesgo de medición.
C. Sesgo de confusión.
D. Azar.
E. Validez externa (capacidad de generalización).

1-3. Los pacientes que fueron asignados al tratamiento quirúrgico eran más jóvenes y gozaban de mejor salud que los que permanecieron con asistencia.

A. Sesgo de selección.
B. Sesgo de medición.
C. Sesgo de confusión.
D. Azar.
E. Validez externa (capacidad de generalización).

1-4. Un menor número de los pacientes que no fueron sometidos a cirugía seguía siendo atendido en la consulta al cabo de 2 meses.

A. Sesgo de selección.

booksmedicos.org

B. Sesgo de medición.
C. Sesgo de confusión.
D. Azar.
E. Validez externa (capacidad de generalización).

1-5. En comparación con los pacientes que únicamente recibieron asistencia médica, puede que los pacientes sometidos a cirugía tuvieran menos probabilidades de notificar cualquier tipo de dolor, y los médicos que los trataban, menos inclinados a registrar el dolor en su historia clínica.

A. Sesgo de selección.
B. Sesgo de medición.
C. Sesgo de confusión.
D. Azar.
E. Validez externa (capacidad de generalización).

1-6. Los pacientes sin otras enfermedades tenían más posibilidades de restablecerse y de ser derivados a una intervención quirúrgica.

A. Sesgo de selección.
B. Sesgo de medición.
C. Sesgo de confusión.
D. Azar.
E. Validez externa (capacidad de generalización).

Seleccione la mejor respuesta para las preguntas 1-7 a 1-11.

1-7. La histamina es un mediador de la inflamación en los pacientes con rinitis alérgica primaveral («fiebre del heno»). De acuerdo con este hecho, ¿cuál de las siguientes afirmaciones es verdadera?

A. Los fármacos que bloquean los efectos de la histamina aliviarán los síntomas.
B. Un descenso de los niveles de histamina en la nariz es un marcador fiable de éxito clínico.
C. Puede que los antihistamínicos resulten eficaces, por lo que sus efectos sobre los síntomas (como prurito nasal, estornudos y congestión nasal) deben estudiarse en los pacientes con rinitis

booksmedicos.org

alérgica.

D. Los otros mediadores no son importantes.

E. Si los estudios analíticos de la enfermedad son convincentes, no es necesaria la investigación clínica

1-8. ¿Cuál de las siguientes afirmaciones sobre las muestras de las poblaciones es incorrecta?

A. Las muestras de poblaciones pueden presentar características que difieran de la población hayan seguido de manera correcta los procedimientos de muestreo.

B. Las muestras de poblaciones son la única forma posible de estudiar la población.

C. Cuando la muestra de la población ha sido correctamente seleccionada, se garantiza la validez externa.

D. Las muestras de poblaciones deberían seleccionarse de modo que cada miembro de la población tenga las mismas posibilidades de ser elegido.

1-9. Al decidir el tratamiento para un paciente de 72 años de edad con cáncer de colon, usted sabe que algunos buenos estudios han demostrado que una determinada combinación de fármacos prolonga la vida de este tipo de pacientes. Sin embargo, los pacientes que participaron en estos estudios eran bastante más jóvenes. ¿Cuál de las siguientes afirmaciones es la correcta?

A. Teniendo en cuenta estos estudios, la decisión de este tratamiento es una cuestión de juicio personal.

B. Confiar o depositar su confianza en estos estudios, se denomina validez interna.

C. Los resultados de estos estudios se ven afectado por el azar, no por el sesgo.

1-10. Se realizó un estudio para determinar si el ejercicio regular disminuye el riesgo de enfermedad cardíaca coronaria (ECC). Se ofreció un programa de ejercicio a los empleados de una fábrica y se compararon las tasas de episodios coronarios subsecuentes en los empleados que se ofrecieron para participar en el programa y en los que no lo hicieron. El

booksmedicos.org

desarrollo de ECC se determinó mediante revisiones voluntarias regulares, incluyendo un interrogatorio cuidadoso, un electrocardiograma y una revisión de los registros rutinarios de salud. Sorprendentemente, los miembros del grupo de ejercicio desarrollaron tasas más elevadas de ECC a pesar de ser los que fumaban menos. ¿Cuál de las siguientes opciones tiene menos probabilidad de explicar este resultado?

A. Los voluntarios tienen mayor riesgo de desarrollar ECC que los que no voluntarios antes que comenzara el estudio.

B. Los voluntarios no aumentaron realmente su ejercicio y la cantidad de este fue la misma en los dos grupos.

C. Los voluntarios recibieron más revisiones y, por lo tanto, fue más factible diagnosticar los infartos silenciosos al miocardio en el grupo que hizo ejercicio.

1-11. Las despolarizaciones ventriculares prematuras se asocian con un riesgo elevado de muerte súbita por una arritmia mortal, sobre todo en personas con otros hallazgos de enfermedad cardíaca. Usted ha leído que existe un nuevo medicamento para tratar las despolarizaciones ventriculares prematuras. ¿Cuál es el dato más importante sobre el medicamento que le gustaría conocer antes de prescribirlo a su paciente?

A. El mecanismo de acción del fármaco.

B. Qué tan bien se previenen las despolarizaciones ventriculares prematuras en personas que tomaron el medicamento en comparación con los que no lo tomaron.

C. La tasa de muerte súbita en personas similares que tomaron o no el medicamento.

Las preguntas 1-12 a 1-15 se basan en el siguiente escenario clínico.

Ya que los informes sugieren que los estrógenos aumentan el riesgo de trombosis, un estudio comparó la frecuencia del uso de anticonceptivos orales entre las mujeres con tromboflebitis ingresadas en un hospital y un grupo de mujeres que ingresaron por otras razones. Se revisaron los registros médicos para conocer la indicación del uso de anticonceptivos orales en los dos grupos. Se observó que las mujeres con tromboflebitis habían tomado anticonceptivos orales con mayor frecuencia que las ingresadas por otras

booksmedicos.org

razones.

Para cada una de las siguientes afirmaciones, seleccione una respuesta que represente un peligro para la validez.

1-12. Es probable que las mujeres con tromboflebitis hayan reportado el uso de anticonceptivos más abiertamente que las mujeres sin tromboflebitis porque recuerdan haber oído algo de la asociación.

A. Sesgo de selección.
B. Sesgo de medición.
C. Sesgo de confusión.
D. Azar.
E. Validez externa (generalización).

1-13. Es probable que los doctores hayan interrogado más cuidadosamente a las mujeres con tromboflebitis sobre el uso de anticonceptivos que a aquellas sin tromboflebitis (y registraron la información con más detalle en el registro médico) porque sabían que los estrógenos podían causar trombosis.

A. Sesgo de selección.
B. Sesgo de medición.
C. Sesgo de confusión.
D. Azar.
E. Validez externa (generalización).

1-14. El número de mujeres incluido en el estudio fue escaso.

A. Sesgo de selección.
B. Sesgo de medición.
C. Sesgo de confusión.
D. Azar.
E. Validez externa (generalización).

1-15. Las mujeres con tromboflebitis fueron admitidas en el hospital por médicos que trabajaban en diferentes barrios que los médicos de aquellas sin tromboflebitis.

booksmedicos.org

A. Sesgo de selección.
B. Sesgo de medición.
C. Sesgo de confusión.
D. Azar.
E. Validez externa (generalización).

Encontrará las respuestas en el apéndice A.

BIBLIOGRAFÍA

1. Home PD, Pocock SJ, Beck-Nielsen H, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomized, open-label trial. Lancet 2009;373:2125–2135.
2. Lipscombe LL, Gomes T, Levesque LE, et al. Thiazolidinediones and cardiovascular outcomes in older patients with diabetes. JAMA 2007;298:2634–2643.
3. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. N Engl J Med 2007;356:2457–2471.
4. Friedman GD. Primer of Epidemiology, 5th ed. New York: Appleton and Lange; 2004.
5. Straus SE, Richardson WS, Glasziou P, et al. Evidence-Based Medicine: How to Practice and Teach EBM, 4th ed. New York: Churchill Livingstone; 2011.
6. Stuebe AM. Level IV evidence—adverse anecdote and clinical practice. N Engl J Med 2011;365(1):8–9.
7. Murphy EA. The Logic of Medicine. Baltimore: Johns Hopkins University Press; 1976.
8. Porta M. A Dictionary of Epidemiology, 5th ed. New York: Oxford University Press; 2008.
9. McCormack K, Scott N, Go PM, et al. Laparoscopic techniques versus open techniques for inguinal hernia repair. Cochrane Database Systematic Review 2003;1:CD001785. Publication History: Edited (no change to conclusions) 8 Oct 2008.
10. Neumayer L, Giobbie-Hurder A, Jonasson O, et al. Open mesh versus laparoscopic mesh repair of inguinal hernia. N Eng J Med 2004;350:1819–1827.
11. Sackett DL. Bias in analytic research. J Chronic Dis 1979;32: 51–63.
12. Pickering TG, Hall JE, Appel LJ, et al. Recommendations for blood pressure in humans and experimental animals. Part 1: Blood pressure measurement in humans. A statement for professionals from the Subcommittee of Professional and Public Education of the American Heart Association Council on High Blood Pressure Research. Circulation 2005;111: 697–716.
13. Bjelakovic G, Nikolova D, Gluud LL, et al. Mortality in randomized trials of antioxidant supplements for primary and secondary prevention: systematic review and meta-analysis. JAMA 2007; 297(8):842–857.
14. Vevekananthan DP, Penn MS, Sapp SK, et al. Use of antioxidant vitamins for the prevention of cardiovascular disease: meta-analysis of randomized trials. Lancet 2003;361: 2017–2023.
15. Norman RJ, Nisenblat V. The effects of caffeine on fertility and on pregnancy outcomes. In: Basow DS, ed. UpToDate. Waltham, MA: UpToDate; 2011.
16. Savitz DA, Chan RL, Herring AH, et al. Caffeine and miscarriage risk. Epidemiology 2008;19:55–62.
17. Sullivan PF. Mortality in anorexia nervosa. Am J Psychiatry 1995;152:1073–1074.
18. Korndorfer SR, Lucas AR, Suman VJ, et al. Long-term survival of patients with anorexia nervosa:

booksmedicos.org

a population-based study in Rochester, Minn. Mayo Clin Proc 2003;78: 278–284.

booksmedicos.org

Capítulo 2

Frecuencia

Aquí, es necesario contar.

—P.C.A. Louis[†]
1787-1872

PALABRAS CLAVE

Numerador
Denominador
Prevalencia
Prevalencia puntual
Prevalencia de período
Incidencia
Duración de la enfermedad
Tasa de letalidad
Tasa de supervivencia
Tasa de complicaciones
Tasa de mortalidad infantil
Tasa de mortalidad perinatal
Estudios de prevalencia
Estudios transversales
Encuestas
Cohorte
Estudios de cohorte
Incidencia acumulada
Densidad de incidencia
Tiempo-persona

booksmedicos.org

Población dinámica
Población en riesgo
Muestra aleatoria
Muestra probabilística
Fracción de muestreo
Sobrerrepresentación
Muestras de conveniencia
Muestras disponibles
Epidemia
Pandemia
Curva epidémica
Endémico

En el capítulo 1 se esbozan las preguntas que los médicos necesitan responder cuando atienden a los pacientes. Las respuestas suelen presentarse en forma de probabilidades y, con muy poca frecuencia, de certezas. Las frecuencias obtenidas en la investigación clínica constituyen la base de las estimaciones de probabilidad para el propósito de la atención de los pacientes. En este capítulo se describen expresiones de frecuencia básicas, la forma en que se obtienen de la investigación clínica y el modo de reconocer las amenazas a su validez.

Ejemplo

Un hombre de 72 años presenta polaquiuria, disuria inicial y goteo posmiccional de progresión lenta. El tacto rectal revela una próstata con un aumento de tamaño simétrico. Las determinaciones del flujo urinario muestran una reducción de la velocidad del mismo, y el antígeno prostático específico (PSA) sérico no está elevado. El médico diagnostica hiperplasia prostática benigna (HPB). Para decidir el tratamiento, médico y paciente deben sopesar los beneficios y los riesgos de las diversas opciones terapéuticas. Para simplificar, podemos decir que las opciones son el tratamiento farmacológico o quirúrgico. El paciente puede elegir el tratamiento con fármacos, pero corre el riesgo de que los síntomas empeoren o de sufrir una nefropatía obstructiva, debido a que la eficacia de este tratamiento es menos inmediata que la de la cirugía. Puede que el paciente opte por la cirugía, con lo que obtendrá un alivio inmediato de los síntomas, pero se expondrá al riesgo de mortalidad operatoria e incontinencia urinaria de larga evolución e impotencia.

Las decisiones como a la que se enfrentan este paciente y médico tradicionalmente han dependido del juicio clínico de acuerdo con la experiencia adquirida en la cabecera del paciente y en las consultas. En los últimos tiempos, la investigación clínica se ha extendido y consolidado lo

booksmedicos.org

bastante para que también se pueda sustentar el juicio clínico en las probabilidades basadas en la investigación (o frecuencias). Las probabilidades de enfermedad, mejora, deterioro, curación, efectos secundarios y muerte constituyen la base para responder la mayor parte de las preguntas clínicas. Para este paciente, la toma de decisiones sólida requiere estimaciones precisas sobre la forma en que cambiarán los síntomas y las complicaciones del tratamiento a lo largo del tiempo según el tratamiento elegido.

[†]Médico del siglo XIX y defensor del «método numérico» (basándose en las cuentas y no en las impresiones) para entender la historia natural de las enfermedades como la fiebre tifoidea.

¿PUEDEN LAS PALABRAS SUSTITUIR ADECUADAMENTE A LOS NÚMEROS?

A menudo, los médicos expresan las probabilidades en forma de palabras (p. ej., habitualmente, a veces, rara vez) en lugar de hacerlo en forma de números. Sustituir los números por palabras es práctico y evita tener que afirmar algo de forma precisa cuando una probabilidad es incierta. Sin embargo, las palabras son pobres sustitutas de los números, debido a que hay poca unanimidad respecto a los significados de los adjetivos que se usan más a menudo para describir la probabilidad.

Ejemplo

Se pidió a los médicos que asignaran valores en porcentaje a 13 expresiones de probabilidad (1). Estos médicos generalmente estaban de acuerdo con las probabilidades que correspondían a adjetivos como «siempre» o «nunca» para describir episodios muy probables o muy poco probables, no así en expresiones asociadas a probabilidades menos extremas. Por ejemplo, el rango de probabilidades (desde el límite superior hasta el décimo inferior de los médicos en servicio) fue del 60-90 % para «usualmente», del 5-45 % para «algunas veces» y del 1-30 % para «rara vez». Esto sugiere (como antes habían afirmado los autores de un estudio) que «la diferencia de opinión entre los médicos, por lo que al tratamiento de un problema se refiere, puede reflejar diferencias en el significado atribuido a las palabras utilizadas para definir la probabilidad» (2).

Los pacientes también asignan una amplia gama de probabilidades a aquello que describen las palabras. En otro estudio, trabajadores profesionales y altamente cualificados, externos al ámbito médico, creyeron

booksmedicos.org

que «habitualmente» correspondía a una probabilidad de entre 35 % y 100 %, mientras que «rara vez» implicaba una probabilidad de entre 0 % y 15 % (3).

Así pues, sustituir los números por palabras reduce la cantidad de información que se expresa, y nosotros recomendamos el empleo de los primeros siempre que sea posible.

PREVALENCIA E INCIDENCIA

En general, las medidas clínicamente relevantes de la frecuencia se expresan como proporciones en las que el **numerador** representa el número de pacientes que presenta un episodio (casos), y el **denominador,** el número de personas en las que podría haber ocurrido dicho episodio (población). Las dos medidas básicas de la frecuencia de morbilidad son la prevalencia y la incidencia.

Prevalencia

La **prevalencia** es la fracción (proporción o porcentaje) de un grupo de personas que presenta una afección clínica o un episodio en un momento determinado. La prevalencia se mide mediante el estudio de una población definida y el recuento del número de personas con y sin la afección objeto de interés. La **prevalencia puntual** se mide en un solo momento del tiempo para cada paciente (si bien, las mediciones reales no tienen por qué realizarse necesariamente en el mismo momento del tiempo calendárico para todas las personas que forman parte de la población). La **prevalencia de período** o **lápsica** describe los casos que existirían en cualquier momento durante un período específico de tiempo.

Incidencia

La **incidencia** es la fracción o proporción de un grupo de personas que al principio no sufrían el episodio de interés, pero que presentan la afección a lo largo de un período determinado. Por lo tanto, la incidencia se refiere a los casos *nuevos* de enfermedad que se producen en una población que inicialmente no sufría la enfermedad o a los nuevos episodios como síntomas o complicaciones que ocurren en pacientes con la enfermedad quienes inicialmente no presentaban estos problemas.

booksmedicos.org

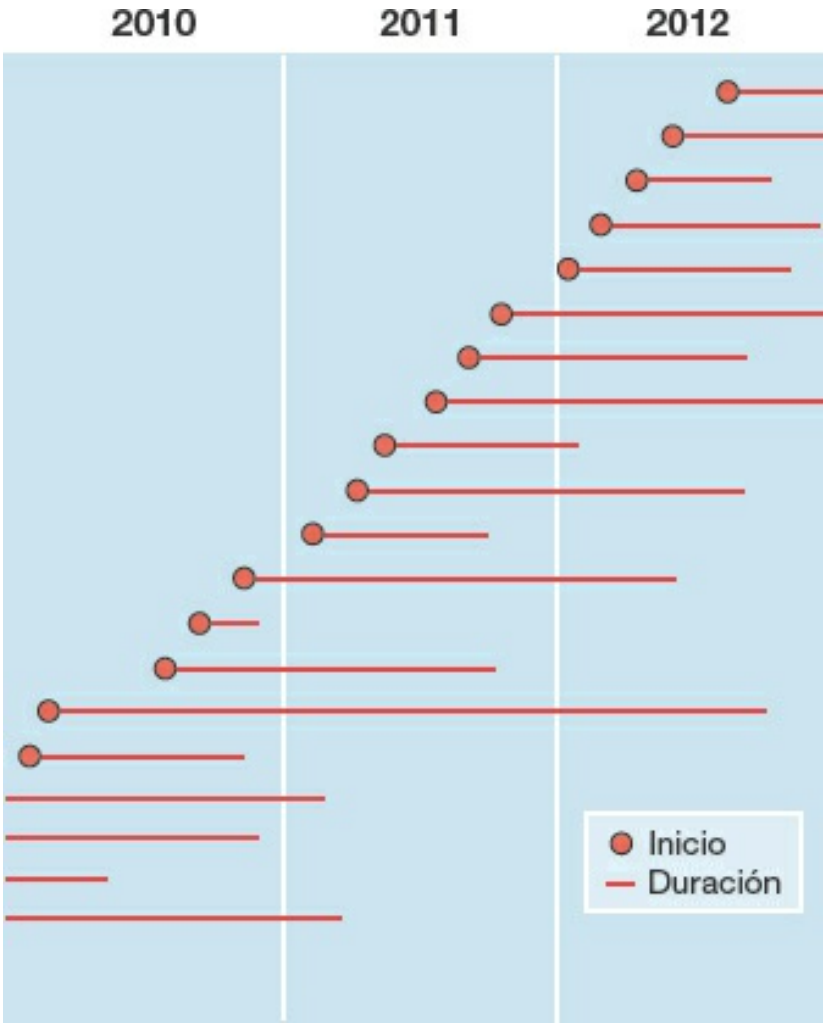


Figura 2-1 ■ Incidencia y prevalencia. Aparición de una enfermedad en 10 000 personas en situación de riesgo entre los años 2010 y 2012.

La figura 2-1 ilustra las diferencias entre la incidencia y la prevalencia. Se muestra la aparición de cáncer de pulmón en una población de 10 000 personas a lo largo de 3 años (2010-2012). Conforme pasa el tiempo, algunas personas que forman parte de la población presentan la enfermedad y permanecen en este estado hasta que se restablecen o mueren. En el caso del cáncer de pulmón, habitualmente mueren. Cuatro personas ya tenían cáncer de pulmón antes de 2010, mientras que 16 lo desarrollaron durante los 3 años de observación. El resto de las 10 000 personas originales no presentaron cáncer de pulmón durante ese período, por lo que no aparecen en la figura.

Para calcular la prevalencia del cáncer de pulmón a principios de 2010, ya existían cuatro casos, por lo que la prevalencia en ese punto es de 4/10 000. Si examinamos a todas las personas supervivientes al inicio de cada año,

booksmedicos.org

podemos calcular la prevalencia en dichos puntos del tiempo. A principios de 2011, la prevalencia es de 5/9 996 porque dos de los pacientes de 2010 siguen vivos, al igual que otras tres personas que desarrollaron la enfermedad en 2010; el denominador se reduce por los 4 pacientes que murieron antes de 2011. La prevalencia se puede calcular para cada uno de los dos exámenes anuales y es de 7/9 992 a inicios de 2011 y de 5/9 986 a inicios de 2012.

Para calcular la incidencia de los nuevos casos que se presentan en la población, nos fijaremos únicamente en las 9 996 personas que a principios de 2010 no sufren la enfermedad y lo que les ocurre a lo largo de los próximos 3 años. En 2010 aparecen cinco casos nuevos de cáncer de pulmón; en 2011, seis, y en 2012, otros cinco casos nuevos. La incidencia de la enfermedad en 3 años está formada por todos los nuevos casos que se han desarrollado a lo largo de estos 3 años (16 casos) divididos por el número de personas susceptibles al principio del período de seguimiento (9 996 personas), o 16/9 996 en 3 años. ¿Cuáles son las incidencias anuales para 2010, 2011 y 2012, respectivamente? Teniendo en cuenta que debemos descontar los casos previos del denominador (pues ya no están en riesgo de desarrollar cáncer de pulmón), el cálculo de las incidencias anuales nos daría 5/9 996 en 2010, 6/9 991 en 2011 y 5/9 985 en 2012.

Prevalencia e incidencia con respecto al tiempo

Cualquier medida de la frecuencia de una enfermedad contiene necesariamente alguna indicación del tiempo. Con medidas de prevalencia se supone que el tiempo es instantáneo, como un fotograma de una película. La prevalencia representa la situación de cada paciente en ese punto del tiempo, aunque, en realidad, es posible que para recoger la información de las diversas personas que conforman la población se hayan necesitado varios meses. Sin embargo, para la incidencia, el tiempo viene definido por el intervalo durante el que se observó si las personas susceptibles presentaban el episodio de interés. En la tabla 2-1 se resumen las características de la incidencia y la prevalencia.

¿Por qué es importante conocer la diferencia entre prevalencia e incidencia? Porque estos conceptos responden a dos preguntas diferentes: por un lado, «¿qué proporción de personas dentro de un grupo padece una afección?», y por el otro, «¿con qué frecuencia surgen nuevos casos en una población definida a medida que pasa el tiempo?». La respuesta a una de las preguntas no puede obtenerse directamente de la respuesta a la otra.

booksmedicos.org

RELACIONES ENTRE PREVALENCIA, INCIDENCIA Y DURACIÓN DE LA ENFERMEDAD

Cualquier factor que aumente la duración de la enfermedad incrementará las probabilidades de que dicho paciente sea detectado en un estudio de prevalencia. Si echamos otro vistazo a la figura 2-1, se confirmará este hecho. Los casos prevalentes son aquellos que continúan afectados en la medida en que los pacientes que se curan, que mueren por causa de la enfermedad o que abandonan la población en estudio dejan de constituir un caso en una encuesta de prevalencia. Como resultado, es más probable que las enfermedades de corta duración se pasen por alto en un estudio de prevalencia. Por ejemplo, el 15 % de todas las muertes por cardiopatía coronaria se producen fuera del hospital en la primera hora del inicio y sin síntomas previos de enfermedad cardíaca. Una prevalencia del estudio podría, por lo tanto, perder casi todos estos episodios y subestimar la verdadera carga de la enfermedad coronaria en la comunidad. Por el contrario, las enfermedades de larga duración están bien representadas en las encuestas de prevalencia, incluso si su incidencia es reducida. La incidencia de la enfermedad inflamatoria intestinal en EE.UU. oscila tan sólo entre 2 y 14 casos por 100 000/año, pero su prevalencia es superior, de 37 a 246/100 000, lo que refleja la naturaleza crónica de la enfermedad (4).

Tabla 2-1

Características de la incidencia y la prevalencia

Características	Incidencia	Prevalencia
Numerador	Nuevos casos que se producen durante un período en un grupo que inicialmente no sufre la enfermedad	Casos existentes en un punto o período
Denominador	Todas las personas susceptibles que no sufren la enfermedad al principio del período	Todas las personas examinadas, incluidos los casos y los que no lo son
Tiempo	Duración del período	Punto o período concreto

booksmedicos.org

Cómo se mide	Estudio de cohortes (v. cap. 5)	Estudio de prevalencia (transversal)

La relación entre la incidencia, la prevalencia y la **duración de la enfermedad** en un estado de equilibrio, en el que ninguna de las variables cambia demasiado a lo largo del tiempo, viene expresada por la siguiente fórmula:

$$\text{Prevalencia} = \text{Incidencia} \times \text{Duración media de la enfermedad}$$

Alternativamente,

$$\text{Prevalencia/Incidencia} = \text{Duración}$$

De forma similar, la prevalencia del cáncer prostático en la autopsia es mucho mayor que su incidencia, pues la mayoría de estos cánceres nunca se volverán lo suficientemente sintomáticos como para ser diagnosticados durante la vida.

Ejemplo

Se midió la incidencia y prevalencia de colitis ulcerosa en el Olmstead County, Minnesota, de 1984 a 1993 (5). La incidencia fue de 8,3/100 000 años-persona y la prevalencia de 229/10 000 personas. La duración promedio de estas enfermedad puede estimarse en 229/100 000 dividido entre 8,3/100 000 = 28 años. Por lo tanto, la colitis ulcerosa es una enfermedad crónica compatible con una larga expectativa de vida. Se cumplió con el supuesto de un estado de equilibrio porque los datos de este mismo estudio mostraron que la incidencia cambió poco durante el intervalo del estudio. Aunque las tasas son diferentes en distintas partes del mundo y están cambiando durante largos períodos, todas reflejan una enfermedad crónica.

OTRAS TASAS

La tabla 2-2 resume algunas tasas utilizadas en atención a la salud. Algunas tasas son expresiones de episodios a lo largo del tiempo. Por ejemplo, la **tasa de letalidad** (o alternativamente, la **tasa de supervivencia**) es la proporción entre el número de personas que sufre una enfermedad y el número de las que fallecen a causa de esta (o que sobreviven a ella). En el caso de enfermedades

booksmedicos.org

agudas como la infección por el virus Ebola, el tiempo de seguimiento puede estar implícito, asumiendo que las muertes se contabilizaron durante un largo período (en este caso, algunas semanas) para considerar todas las que podrían haber ocurrido. En el caso de las enfermedades crónicas, como la enfermedad cardiovascular o el cáncer, es más habitual especificar el período de observación (p. ej., una tasa de supervivencia de 5 años). De forma similar, la **tasa de complicaciones**, la proporción de personas con la enfermedad o en tratamiento, y que presentan complicaciones, supone que ha pasado una cantidad de tiempo suficiente para que ya se hayan producido las complicaciones. Este tipo de medidas puede dar lugar a infravaloraciones si el tiempo de seguimiento no es lo bastante prolongado. Por ejemplo, se notifican menos tasas de infección del sitio de incisión quirúrgica de lo que se debiera, a causa de que se han contabilizado hasta el momento del alta hospitalaria, mientras que la infección de algunas heridas sólo se manifiesta después del alta (6).

Otras tasas, como la **tasa de mortalidad infantil** y la **tasa de mortalidad perinatal** (definidas en la tabla 2-2) constituyen aproximaciones de la incidencia, debido a que los niños que figuran en el numerador no son necesariamente los mismos que los del denominador. En el caso de la mortalidad infantil para un año concreto, algunos de los niños que fallecen en dicho año nacieron en el anterior; de forma similar, el último niño que nació aquel año puede morir al siguiente. Estas tasas se elaboran de esta forma para que la medición resulte más factible, a la vez que proporcionan una aproximación útil de la tasa real en un año determinado.

Tabla 2-2

Tasas utilizadas más habitualmente

Tasa de letalidad	Proporción de pacientes que muere a causa de una enfermedad
Tasa de complicaciones	Proporción de pacientes que sufre una complicación derivada de una enfermedad o de su tratamiento
Tasa de mortalidad infantil	$\frac{\text{Número de muertes durante 1 año en niños < 1 } \textit{año de edad}}{\text{Número de nacidos vivos a lo largo del mismo año}}$
Tasa de mortalidad perinatal (definición de la OMS)	Número de muertes fetales y defunciones ocurridas en la primera semana de vida por cada 1 000 nacidos vivos
Tasa de mortalidad materna	$\frac{\text{Número de muertes maternas relacionadas con el parto en un año determinado}}{\text{Número de nacidos vivos en la misma población durante ese mismo año}}$

booksmedicos.org

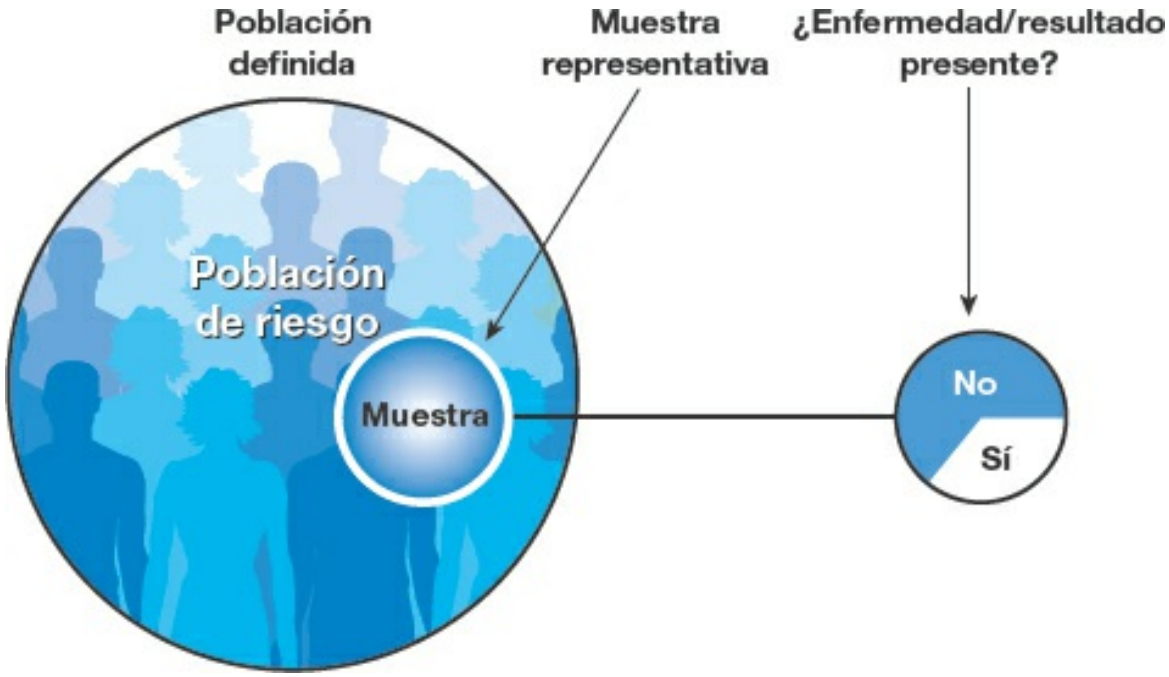


Figura 2-2 ■ Diseño de un estudio de prevalencia.

ESTUDIOS DE PREVALENCIA E INCIDENCIA

Para medir la prevalencia y la incidencia se utilizan tipos de estudios completamente diferentes.

Estudios de prevalencia

En los **estudios de prevalencia** se examina a las personas que forman parte de una población en busca de la afección de interés. En ese punto del tiempo, algunos miembros de la población sufren la enfermedad y otros no (fig. 2-2). La fracción o proporción de la población que tiene la afección (es decir, los casos) constituye la prevalencia de la enfermedad.

Otro término utilizado es el de **estudios transversales**, ya que las personas son estudiadas en un punto (corte transversal) del tiempo. Los estudios de prevalencia también se denominan **encuestas** si la medida principal es un cuestionario.

El ejemplo siguiente ilustra un estudio de prevalencia clásico.

Ejemplo

booksmedicos.org

La Organización Mundial de la Salud creó un consorcio de investigación para estudiar la prevalencia de depresión en varios países. Se entrevistó a más de 37 000 personas, seleccionadas al azar en la población general en 10 países de Norteamérica, Latinoamérica, Europa y Asia. Los episodios de depresión mayor fueron diagnosticados mediante entrevistas cara a cara, utilizando un instrumento diseñado para obtener resultados uniformes en diferentes idiomas y culturas. Las tasas de respuesta oscilaron entre el 57-90 %. La prevalencia a 30 días (de hecho es la prevalencia durante un período, pero es una buena estimación de la prevalencia puntual debido al reducido margen de tiempo) osciló entre un máximo del 4,6 % en EE.UU. y a un mínimo del 0,9 % en Japón. La prevalencia durante el período fue mayor; por ejemplo, en EE.UU. la prevalencia a 12 meses fue del 10 % y la prevalencia a lo largo de la vida del 16,9 %. Los autores concluyeron que «los episodios de depresión mayor son un trastorno habitual que por lo general tienen un curso crónico intermitente» (7).

Estudios de incidencia

La población examinada en un estudio de incidencia se denomina **cohorte,** y se define como un grupo de personas que tienen algún factor en común en el momento en que se forma el grupo, y que son seguidas posteriormente a lo largo de un tiempo a la espera de que se produzcan resultados de interés. Por esta razón, los estudios de incidencia también se denominan **estudios de cohortes.** Se identifica una muestra de personas sin el episodio de interés y se observa a lo largo del tiempo, para detectar si se produce algún resultado. Puede que los miembros de la cohorte gocen de un buen estado de salud al principio y que sean vigilados durante el tiempo para detectar el inicio de la enfermedad, por ejemplo de estar libres de cáncer hasta el inicio (o no) del cáncer pancreático. O bien, todos ellos pueden tener una enfermedad de diagnóstico reciente (como cáncer pancreático) y ser vigilados durante el tiempo para evaluar los resultados como la recurrencia o la muerte. Los estudios de incidencia se abordarán con mayor detalle en los capítulos 5 y 7.

Incidencia acumulada

Hasta el momento, el término «incidencia» se ha utilizado para describir la proporción de nuevos episodios en un grupo de personas con un tamaño fijado, cuyos miembros son observados en su totalidad durante un período de tiempo. Es lo que se denomina **incidencia acumulada,** debido a que los nuevos casos se acumulan a lo largo del tiempo.

Densidad de incidencia (años-persona)

booksmedicos.org

Otro método para estudiar la incidencia consiste en medir el número de casos nuevos que surgen en una población en constante cambio, en la que las personas se encuentran sometidas a estudio y son susceptibles durante períodos de tiempo variables. La incidencia derivada de este tipo de estudios se denomina **densidad de incidencia,** puesto que constituye, en términos figurados, la densidad de nuevos casos en un período de tiempo y en un lugar concretos.

Los ensayos clínicos a menudo utilizan el abordaje de densidad de incidencia. Los pacientes elegibles se someten a estudio durante un tiempo, de forma que los pacientes incluidos al principio son tratados y seguidos durante períodos de tiempo más prolongados que los pacientes incluidos al final. En un esfuerzo para que la contribución de cada paciente sea proporcional a su intervalo de seguimiento, el denominador de una medida de densidad de incidencia no lo constituyen las personas en riesgo durante un período específico, sino el **tiempo-persona** en situación de riesgo de sufrir el parámetro de interés. Un paciente sometido a seguimiento durante 10 años sin sufrir ninguno de los parámetros de interés contribuye en 10 años-persona, mientras que otro paciente sometido a seguimiento durante 1 año sólo contribuye al denominador en 1 año-persona. La densidad de incidencia se expresa como el número de casos nuevos en relación con el número total de años-persona en riesgo.

El método de años-persona es especialmente útil para calcular la incidencia de la enfermedad en **poblaciones dinámicas,** donde algunos individuos están entrando y otros están abandonando la población conforme pasa el tiempo. Los estudios de incidencia en poblaciones grandes suelen tener un recuento preciso de los casos nuevos dentro de la población (p. ej., de archivos hospitalarios o registros sobre la enfermedad), pero el tamaño y las características de la población en riesgo sólo pueden ser estimadas (de censos y otros registros) porque las personas están entrando y abandonando la región continuamente. Este abordaje funciona porque la proporción de personas que entran o que abandonan es pequeña en comparación con la totalidad de la población (fig. 2-3), por lo que es probable que la población se mantenga relativamente estable durante períodos cortos.

Ejemplo

booksmedicos.org

Un estudio sobre la incidencia de infecciones por herpes zóster («culebrilla») y sus complicaciones proporciona un ejemplo de la densidad de incidencia y de la incidencia acumulada. Investigadores del Olmstead County, Minnesota, revisaron los archivos médicos de los residentes mayores del condado (8). Otros estudios mostraron que más del 98 % de toda la atención prestada a los residentes se proporcionó dentro del condado y la mayoría de los residentes estuvieron de acuerdo en permitir que sus archivos médicos se utilizaran para la investigación. A partir de la información del censo se estimó la población del condado en 175 000 personas. La incidencia de herpes zóster, ajustada a la edad y al género de la población adulta de EE.UU., fue de 3,6 por 1 000 años-persona y aumentó con la edad. El 18 % de estos pacientes presentaron dolor tras un ataque de herpes.

La incidencia de la infección por herpes zóster se describió en años-persona en una población dinámica, mientras que el dolor después de la infección fue una incidencia acumulada, en la que se siguió a todos los pacientes con herpes zóster.

Un inconveniente del método años-persona es que agrupa períodos de seguimiento de duraciones diferentes. Un reducido número de pacientes controlados durante un largo período puede contribuir en los mismos años-persona que un gran número de enfermos seguidos durante un breve período. Si los pacientes sometidos a un largo seguimiento son sistemáticamente diferentes de los que se someten a uno breve (quizá porque los resultados de interés tardan más en aparecer o porque los pacientes con unas perspectivas especialmente malas tienden a abandonar la población), la densidad de incidencia resultante dependerá de la combinación concreta del número de pacientes y el tiempo de seguimiento. Por ejemplo, el período de latencia entre la exposición al carcinógeno y el inicio del cáncer es de al menos 10 años en la mayoría de los cánceres. Podría observarse un aumento de la tasa de cáncer en un estudio que incluyera a 1 000 personas expuestas a un carcinógeno y sometidas a seguimiento durante 20 años. Pero no sucedería lo mismo con un estudio de 10 000 personas sometidas a seguimiento durante 2 años, a pesar de que el número de años-persona (200 000) resultante es el mismo, puesto que el período de seguimiento es demasiado breve.

booksmedicos.org

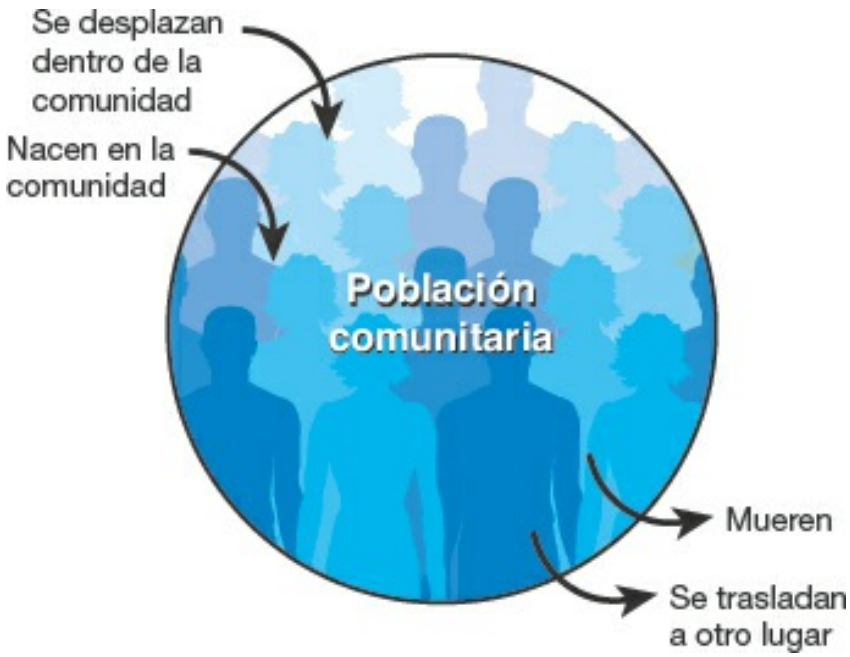


Figura 2-3 ■ Población dinámica.

ELEMENTOS BÁSICOS DE LAS MEDIDAS DE FRECUENCIA

Para comprender bien una tasa de incidencia o de prevalencia, ante todo es necesario definir minuciosamente los términos numerador y denominador.

¿Qué es un caso?: definición de numerador

Los casos pueden ser personas de la población general que presentan una enfermedad o pacientes de ámbitos clínicos que presentan un resultado de interés como la recurrencia, la complicación de un tratamiento o la muerte. Cualquiera que sea la situación, la forma en que se define un caso influye en las tasas.

La mayoría de los datos clínicos (colesterol sérico, calcio sérico, niveles de hormonas tiroideas, etc.) son continuos, de niveles bajos a elevados. El punto de corte para definir un caso puede establecerse en distintos valores y esto puede tener un efecto importante sobre la prevalencia resultante. En el capítulo 3 se discutirán algunas de las razones por las que podríamos elegir uno u otro punto de corte y en el capítulo 8 se comentarán las consecuencias para el rendimiento de una prueba diagnóstica.

booksmedicos.org

Ejemplo

Hay una epidemia de obesidad en el mundo. ¿Cuál es la prevalencia de peso anormal en EE.UU.? Depende de cómo se defina anormal. La figura 2-4 muestra la distribución del índice de masa corporal (IMC, una forma de medir la obesidad que tiene en cuenta el peso y la talla) en los hombres y las mujeres estadounidenses de 40 a 59 años de 2007 a 2008 (9). Los U.S. National Institutes of Health y la Organización Mundial de la Salud han recomendado una clasificación del IMC (tabla 2-3). De acuerdo con esta clasificación, la prevalencia de la obesidad fue del 33,8 %, mientras que la prevalencia del sobrepeso (que incluye personas con obesidad y más) del 68 %. Una proporción significativa de la población, cerca del 5 %, estaba en el grado más extremo (grado III, «obesidad mórbida») de peso.

Las tasas también pueden verse influidas por el grado de intensidad con que el médico busca los casos. Por ejemplo, el ácido acetilsalicílico puede inducir asma en algunas personas. ¿Con qué frecuencia se produce este fenómeno? Depende de la definición de un caso. Cuando se pregunta a las personas sencillamente si tienen un problema respiratorio después de tomar ácido acetilsalicílico, las tasas son relativamente bajas, de un 3 % en los adultos. Cuando el caso se define de forma más rigurosa, se administra ácido acetilsalicílico y se mide si a continuación se produce broncoconstricción, la prevalencia del asma inducida por el ácido acetilsalicílico es mucho mayor, un 21 % en los adultos (10). La menor de ambas tasas está relacionada con las situaciones clínicas, mientras que la mayor aporta información sobre la biología de esta enfermedad.

booksmedicos.org

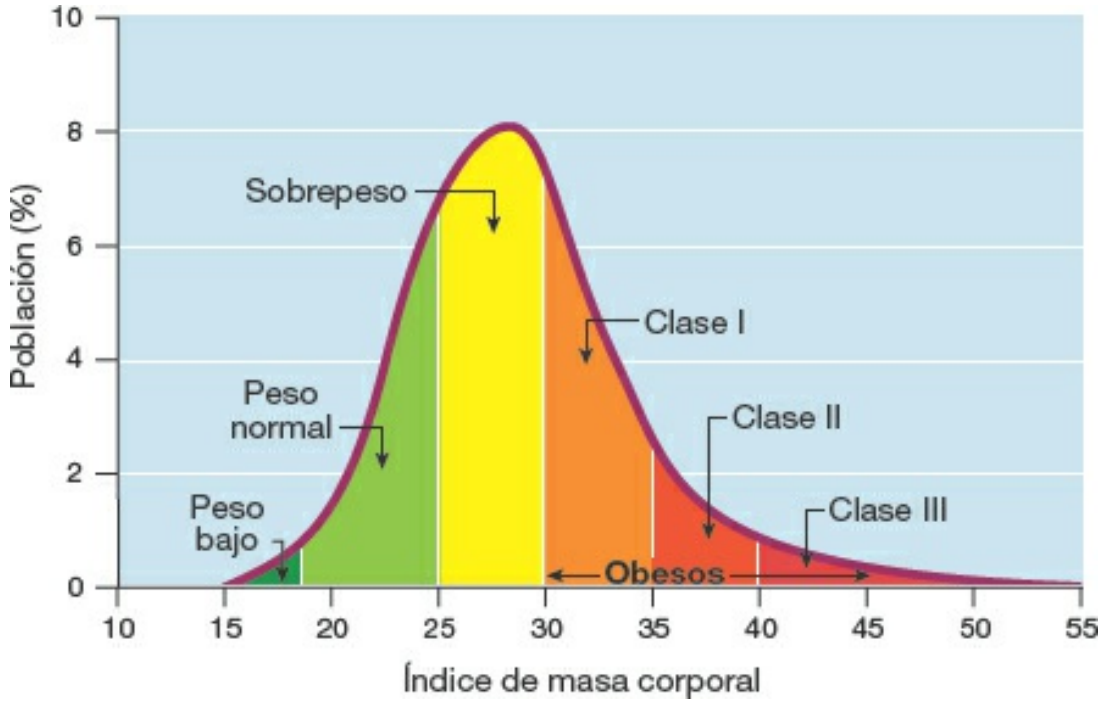


Figura 2-4 ■ La prevalencia del sobrepeso y la obesidad en los hombres de 2007 a 2008. (Datos de Flegal KM, Carroll MD, Ogden CL, et al. Prevalence and trends in obesity among US adults, 1999-2008. JAMA 2010;303[3]:235-241.)

Tabla 2-3

Clasificación de la obesidad de acuerdo con los U.S. National Institutes of Health y la OMS

Clasificación	Índice de masa corporal (kg/m^2)
Peso bajo	<18,5
Peso normal	18,5-24,9
Sobrepeso	25,0-29,9
Obesidad Grado I Grado II Grado III («severa», «extrema» o «mórbida»)	≥30 30,0-34,9 35,0-39,9 ≥40

La incidencia también puede cambiar en función de lo eficaces que sean

booksmedicos.org

los métodos para detectar la enfermedad.

Ejemplo

Muchos casos de cáncer de próstata permanecen asintomáticos y no se detectan a lo largo de la vida. En la actualidad, con la utilización del PSA, un análisis sanguíneo para la detección del cáncer de próstata, se identifican muchos de dichos casos. El análisis es relativamente sensible y guía hacia las biopsias prostáticas que descubren cánceres que, de lo contrario, permanecerían sin detectar. Como resultado del uso extendido de la prueba del PSA, se ha producido un rápido aumento en la incidencia de cáncer de próstata, más del doble en unos pocos años (11) (fig. 2-5). La elevación de la incidencia no se corresponde con un incremento real de la incidencia en la población, ya que este se ha producido de forma muy rápida y no se han observado aumentos similares en otros países, donde la prueba del PSA no es tan habitual. La incidencia ha presentado subsecuentemente un cierto descenso, tal vez porque el reservorio de casos prevalentes, atraídos por esta nueva prueba, se ha agotado. Sin embargo, la incidencia no ha disminuido hasta los niveles previos de la prueba de PSA y parece haber alcanzado una meseta más alta, sugiriendo que se están diagnosticando casos nuevos con mayor frecuencia desde que se introdujo dicha prueba.

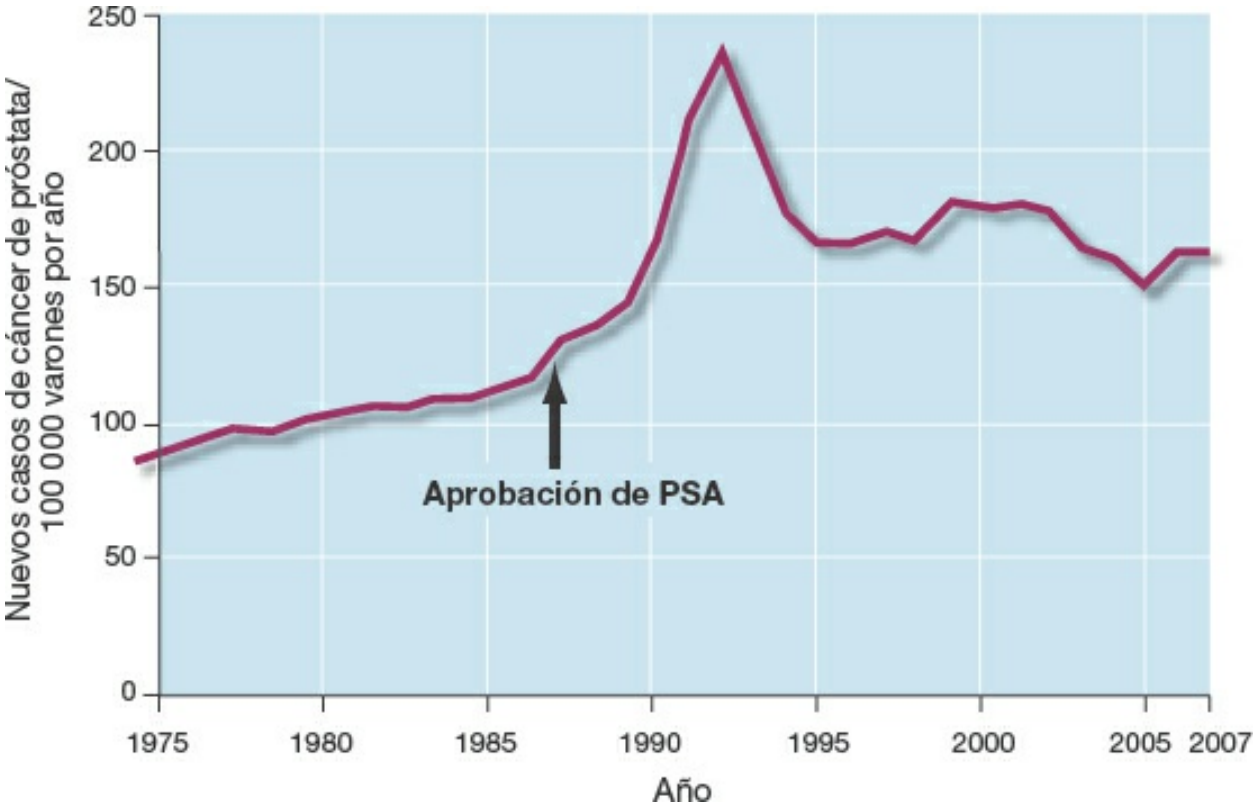


Figura 2-5 La incidencia depende de la intensidad de los esfuerzos para encontrar casos. Incidencia de cáncer de próstata en EE.UU. durante el amplio uso de la detección con antígeno prostático específico (PSA). (Redibujado con autorización de Wolf AMD, Wender RC, Etzioni RB et

booksmedicos.org

al. American Cancer Society guideline for the early detection of prostate cancer: Update 2010. CA Cancer Journal for Clinicians 2010;60:70-98.)

¿Qué es la población?: definición de denominador

Una tasa es útil sólo en la medida en que la población en la que se mide (el denominador de la tasa) se encuentre claramente definida y resulte idónea para la cuestión. Existen tres características del denominador que son especialmente importantes.

En primer lugar, todos los miembros de la población deben poder presentar el resultado de interés, es decir, deben formar una **población en riesgo.** Si los miembros de la población no pueden presentar el episodio o la afección que se cuenta en el numerador, significa que no pertenecen al denominador. Por ejemplo, las tasas de cáncer de cuello uterino deben evaluarse en las mujeres que todavía poseen cuello uterino; en la medida en que los cálculos de las tasas de este cáncer se encuentren en poblaciones que incluyen mujeres histerectomizadas (o para el caso, hombres), las cifras reales estarán subestimadas.

En segundo lugar, la población debe ser adecuada para la pregunta que se formula. Por ejemplo, si deseáramos conocer la prevalencia de la infección por VIH en la comunidad, estudiaríamos una muestra al azar en las personas de la región. Pero si quisiéramos conocer la prevalencia de la infección por VIH entre las personas que toman drogas ilegales, los estudiaríamos a ellos.

En tercer lugar, la población debe describirse con suficiente detalle para crear una base útil para juzgar a quiénes se aplican los resultados del estudio de prevalencia. Esto hace referencia a la capacidad de generalización de las tasas, es decir, la decisión de si una tasa notificada es aplicable al tipo de pacientes del ámbito de interés. Existe un gradiente enorme en las tasas de enfermedad (p. ej., infección por VIH) entre los ámbitos de las diferentes consultas, desde la población general hasta el ámbito de atención primaria, pasando por los centros de derivación. Para poder utilizar la información de forma eficaz, los médicos necesitan localizar las tasas notificadas en ese espectro.

¿La muestra del estudio representa a la población?

Tal como mencionamos en el capítulo 1, en muy pocas ocasiones es posible estudiar a todas las personas que sufren o pueden presentar la afección de interés. Habitualmente se toma una muestra a fin de que el número de

booksmedicos.org

personas estudiadas tenga un tamaño manejable, lo que conduce a una pregunta clave: ¿la muestra representa con exactitud a la población de origen?

Las muestras aleatorias pretenden producir muestras representativas de la población. En una **muestra aleatoria** simple cada miembro de la población tiene la misma probabilidad de ser seleccionado. Se utiliza un término más general, **muestra probabilística,** cuando cada persona tiene una probabilidad conocida (no necesariamente idéntica) de ser seleccionada. La utilidad de las muestras probabilísticas radica en que a menudo la inclusión en la muestra de un número suficiente de personas pertenecientes a subgrupos concretos de interés, como minorías étnicas o pacientes ancianos, aporta mayor cantidad de información. Si los miembros de estos subgrupos constituyen sólo una pequeña proporción de la población, una simple muestra aleatoria de toda la población podría no incluirlos en número suficiente. Para solucionar esto, los investigadores pueden modificar la **fracción de muestreo**, es decir, la fracción de todos los miembros de cada subgrupo incluido en la muestra. También pueden **sobrerrepresentar** (sobremuestrear) a los grupos de baja frecuencia en relación con el resto, es decir, seleccionar al azar una fracción mayor de ellos. La muestra final seguirá siendo representativa de toda la población si, en el análisis, se tienen en cuenta las diferentes fracciones de muestreo.

Por término medio, las características de las personas incluidas en las muestras probabilísticas son parecidas a las de la población de la que fueron seleccionadas, especialmente si el tamaño de la muestra es considerable. La medida en que la muestra difiere de la población de origen se debe al azar y no al error sistemático.

Las muestras no aleatorias son habituales en la investigación clínica por razones prácticas. Se conocen como **muestras de conveniencia** (porque su virtud principal es que se obtuvieron por comodidad, por ejemplo, muestras de pacientes que se atienden en un centro médico, que cooperan y que se expresan con claridad) o **muestras disponibles** (porque los investigadores se limitaron a incluir a los pacientes allí donde los encontraron).

La mayoría de los pacientes que se describen en la bibliografía médica y los atendidos por los médicos son muestras sesgadas de la población principal. Típicamente, los pacientes se incluyen en los estudios porque son atendidos en una institución académica, están disponibles, desean ser estudiados, no están afectados por otras enfermedades aparte de la

booksmedicos.org

enfermedad objeto de estudio y puede que incluso sean especialmente interesantes, estén gravemente afectados, o ambas cosas. No hay nada erróneo en esta práctica, siempre que se entienda a quiénes se pueden (o no se pueden) aplicar los resultados. Sin embargo, debido a las muestras sesgadas, los resultados de la investigación clínica a menudo dejan a los médicos pensativos con un gran problema de generalización, desde el ámbito de la investigación hasta la consulta.

DISTRIBUCIÓN DE LA ENFERMEDAD SEGÚN EL TIEMPO, EL LUGAR Y LAS PERSONAS

La epidemiología se ha descrito como el estudio de los determinantes de la distribución de enfermedad en las poblaciones. Los determinantes principales son el tiempo, el lugar y las personas. La distribución según estos factores puede proporcionar sólidos indicios respecto a las causas y el control de la enfermedad, así como a la necesidad de servicios sanitarios.

Tiempo

Una **epidemia** es una concentración de nuevos casos en el tiempo. El término **pandemia** se utiliza cuando una enfermedad se encuentra bastante extendida, como una epidemia global de influenza particularmente grave (p. ej., la de 1918-1919) o el aumento en todo el mundo, aunque con una evolución más lenta, de la infección por el VIH/SIDA. La existencia de una epidemia se detecta mediante una **curva epidémica,** que muestra el aumento (y, en ocasiones, el descenso) de casos de una enfermedad a lo largo del tiempo en una población.

Ejemplo

La figura 2-6 muestra la curva epidémica de una enfermedad detectada recientemente denominada síndrome respiratorio agudo severo (SARS) en Pekín, República Popular de China, durante la primavera de 2003 (12). En total, se notificaron 2 521 casos probables durante la epidemia. Los casos se clasificaron como «probables» porque, en su momento, no había criterios estrictos para establecer un diagnóstico. La definición de caso empleada incluía combinaciones de los siguientes fenómenos epidemiológicos y clínicos: contacto con un paciente con SARS o el hecho de vivir o visitar una región en la que el SARS estuviera activo, síntomas y signos de enfermedad respiratoria febril, cambios en la radiografía torácica, falta de respuesta a los antibióticos y recuento de leucocitos normal o disminuido. Posteriormente, cuando se obtuvo más información

booksmedicos.org

sobre esta enfermedad, pudo utilizarse el análisis de laboratorio del coronavirus responsable para definir un caso. Los casos pasaron a clasificarse como «notificados» para dejar claro que no había garantías de que se hubieran detectado todos los casos en la comunidad de Beijing.

La figura 2-6 también muestra el momento en que se establecieron importantes medidas de control. La epidemia disminuyó al tiempo que se aplicaron intensas medidas de cuarentena, que conllevaron el cierre de lugares de uso público, la identificación de nuevos casos al inicio de su evolución, la eliminación de los casos de la comunidad y el aislamiento de los casos en instalaciones específicas para el SARS. Es posible que la epidemia remitiera por otras razones además de estas medidas de control, si bien esta opción parece poco probable, ya que medidas de control similares en otros lugares también fueron seguidos por una resolución de la epidemia. Cualquiera que fuera la causa, la disminución de nuevos casos permitió que la Organización Mundial de la Salud levantara su recomendación de no viajar a Beijing, de forma que la ciudad pudo retomar sus actividades comerciales internacionales y reabrir los lugares públicos.

El conocimiento de una epidemia local ayuda a los médicos a llegar al diagnóstico correcto. Por ejemplo, mientras trabajaban como médicos de atención primaria en una base militar de Alemania, uno de ellos vio a un niño con fiebre y exantema en manos y pies, y se quedaron perplejos, ya que la única experiencia que tenían en pediatría era la adquirida durante las prácticas de medicina en un hospital. Sin embargo, cuando empezaron a ver a numerosos niños con los mismos signos en un breve período de tiempo, se dieron cuenta (con la ayuda de un amigo pediatra) de que se encontraban en medio de un brote de infección por virus Coxsackie («enfermedad mano-pie-boca»), una enfermedad infecciosa leve característica de los niños.

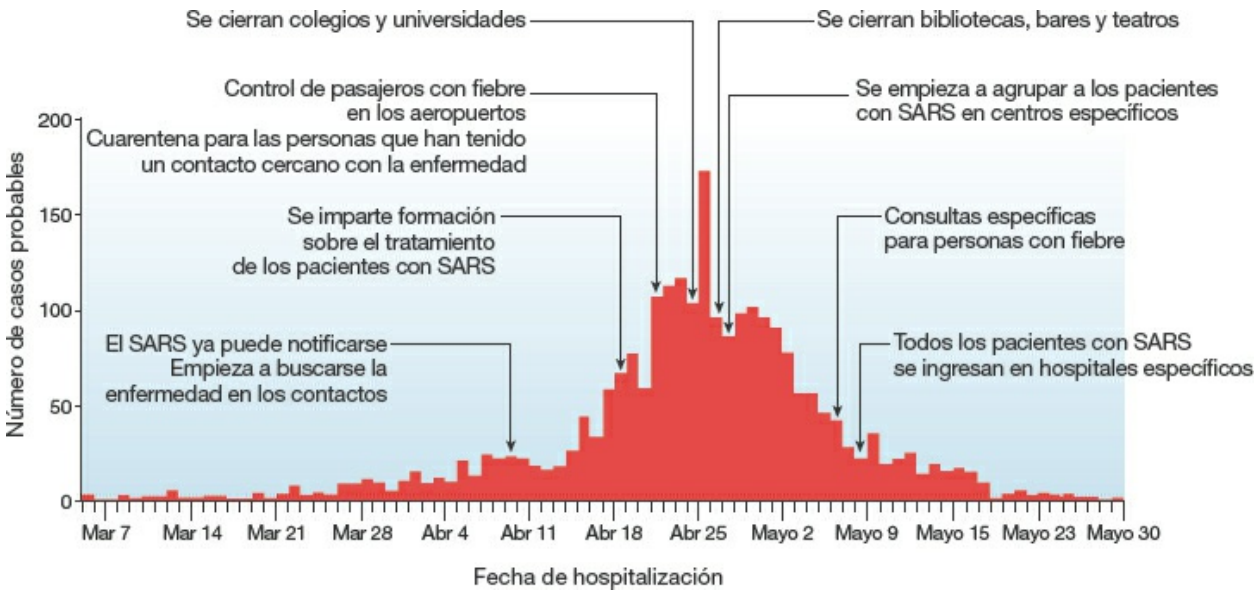


Figura 2-6 ■ Curva epidémica. Casos probables de síndrome respiratorio agudo severo (SARS) en Pekín, marzo-mayo de 2003, en relación con las medidas de control. (Adaptado con autorización de

booksmedicos.org

Pang X, Zhu Z, Xu F, et al. Evaluation of control measures implemented in the severe acute respiratory syndrome outbreak in Beijing, 2003. JAMA 2003;290:3215-3221.)

Lugar

La distribución geográfica de los casos indica los puntos en que la enfermedad es más y menos importante, y aporta indicios sobre sus causas.

Ejemplo

La incidencia del cáncer colorrectal es muy diversa en diferentes partes del mundo. Las tasas, incluso una vez ajustadas por diferencias en cuanto a edad, son elevadas en EE.UU., Europa y Australia, y bajas en África y Asia (fig. 2-7) (13). Esta observación ha dado lugar a la hipótesis de que los factores ambientales pueden desempeñar un papel importante en el desarrollo de la enfermedad. Esta hipótesis ha sido respaldada por otros estudios que muestran que las personas que proceden de países con una incidencia baja y se trasladan a países con una incidencia elevada adquieren tasas de cáncer colorrectal más altas a lo largo de su vida.

Cuando una enfermedad como el bocio por deficiencia de yodo o la poliomielitis (después de los esfuerzos globales para erradicarla) se limitan a ciertas áreas, se denomina **enfermedad endémica.**

Personas

El hecho de que una enfermedad afecte a ciertos tipos de personas en el mismo período y lugar en los que otras personas no se ven afectadas aporta indicios sobre sus causas, así como una orientación sobre cómo desplegar los recursos sanitarios. Al principio de la pandemia de SIDA, la mayor parte de los casos se observaban entre varones homosexuales con múltiples parejas sexuales y entre usuarios de drogas intravenosas. Este hecho dio lugar a la hipótesis inicial de que la enfermedad estaba causada por un agente infeccioso que se transmitía a través del semen y la sangre. Los estudios analíticos confirmaron esta hipótesis e identificaron el VIH. La identificación de los tipos de personas más afectadas también desembocó en esfuerzos especiales para evitar la diseminación de la enfermedad entre ellas, por ejemplo, mediante la formación sobre los métodos de seguridad en las relaciones sexuales dirigida a dichos colectivos, el cierre de los baños públicos y el establecimiento de programas de seguridad sobre la utilización de jeringuillas.

booksmedicos.org

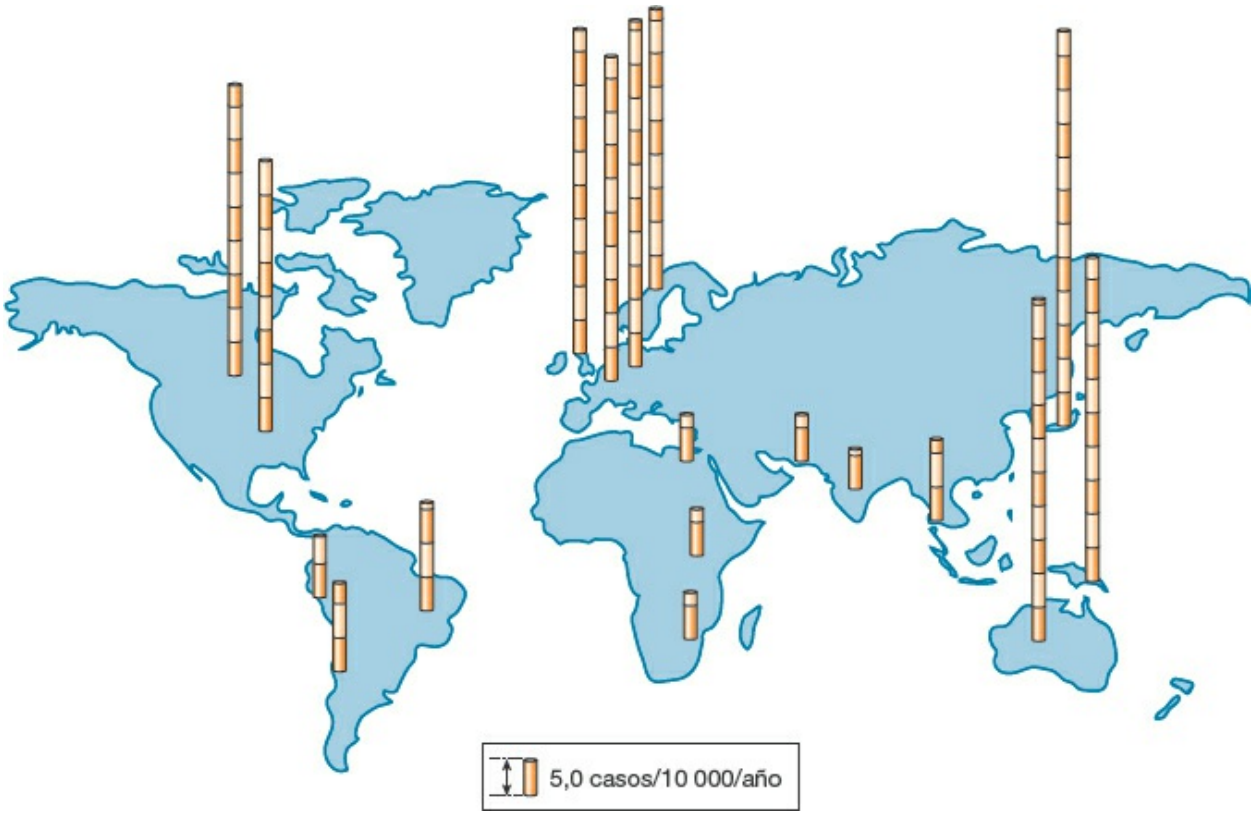


Figura 2-7 ■ Incidencia de cáncer colorrectal en varones en función de la zona del planeta. (Datos de Center MM, Jemal A, Smith RA, et al. Worldwide variations in colorectal cancer. CA Cancer J Clin 2009;59:366-378.)

USOS DE LOS ESTUDIOS DE PREVALENCIA

Los estudios de prevalencia son la mejor manera de responder a algunas preguntas importantes y un método insuficiente para responder a otras.

¿Cuál es la utilidad de los estudios de prevalencia?

Los estudios de prevalencia proporcionan información valiosa sobre aquello que se puede esperar en diferentes situaciones clínicas.

Ejemplo

El tratamiento de la linfadenopatía cervical depende del lugar y del paciente en el que se observe. Los niños con adenopatía cervical persistente que acuden a las consultas de atención primaria sólo tienen un 0,4 % de probabilidades de que el ganglio represente un cáncer, sobre todo un linfoma. Los médicos deben pensar cuidadosamente en solicitar una biopsia, el método definitivo para determinar si está o no presente el cáncer. Sin embargo, los adultos que acuden a las consultas de

booksmedicos.org

atención primaria tienen un 4 % de probabilidades de sufrir un cáncer subyacente de cabeza y cuello. En estos casos, los médicos deben contar con un bajo umbral a la hora de indicar una biopsia de nódulos linfáticos. Las tasas de neoplasias malignas en los centros de referencia son mucho mayores, de un 60 % en los adultos con adenopatía cervical, y habitualmente se realizan biopsias. La situación cambia en las diferentes partes del mundo. En los países en vías de desarrollo, las infecciones por micobacterias representan una causa de linfadenopatía más común que el cáncer. Por consiguiente, conocer la prevalencia contribuye a que los médicos prioricen las posibilidades diagnósticas en su ámbito concreto y para el paciente que atienden en ese momento.

La prevalencia de la enfermedad afecta poderosamente a la interpretación de los resultados de las pruebas diagnósticas, como se describirá con mayor detalle en el capítulo 8.

Finalmente, la prevalencia constituye una importante guía a la hora de planificar los servicios sanitarios. En el ámbito de la atención primaria, la preparación para la diabetes, la obesidad, la hipertensión arterial y los trastornos lipídicos, exige más atención que la planificación para la enfermedad de Hodgkin, la anemia aplásica o el lupus eritematoso sistémico. Por otro lado, algunos hospitales de referencia están bien preparados para estas enfermedades, y de forma adecuada.

¿Para qué no son particularmente buenos los estudios de prevalencia?

Los estudios de prevalencia sólo proporcionan una débil evidencia sobre la causa y el efecto. Las preguntas causales se enfocan esencialmente hacia los nuevos episodios que surgen con el tiempo, es decir, son preguntas relacionadas con la incidencia. Una de las limitaciones de los estudios de prevalencia es que puede ser difícil determinar si la supuesta causa realmente precedió al efecto, ya que ambos se miden en el mismo punto del tiempo. Por ejemplo, si los pacientes hospitalizados con hiperglucemia se infectan con mayor frecuencia, ¿es porque la hiperglucemia altera la función inmunitaria causando la infección o es la infección la que ha provocado la hiperglucemia? Si un factor de riesgo ha precedido al inicio de la enfermedad o al resultado (p. ej., antecedentes familiares o un marcador genético), la interpretación de la secuencia causa-efecto es menos preocupante.

Otra limitación es que la prevalencia puede ser resultado de la incidencia de la enfermedad, la principal consideración en las preguntas causales, o puede estar relacionada con la duración de la enfermedad, un tema completamente diferente. Si sólo se dispone de información acerca de la prevalencia, es imposible determinar cuánto influye cada una, la incidencia y

booksmedicos.org

la duración. Sin embargo, los estudios transversales pueden proporcionar hipótesis convincentes acerca de la causa y el efecto que pueden probarse en estudios más sólidos.

El mensaje fundamental es que un estudio transversal bien realizado, o cualquier otro estudio de investigación, no es sólido o débil por sí mismo sino en relación con la pregunta que intenta responder.

Ejemplo

Los niños que viven en granjas tienen menor probabilidad de tener asma que los que viven en la misma región fuera de una granja. El entorno de la granja y la ciudad difieren en la exposición a microorganismos, sugiriendo que la exposición a estos puede proteger contra el asma. Pero estos entornos difieren de muchas otras maneras. Para fortalecer esta hipótesis, investigadores alemanes examinaron el polvo del colchón y también el polvo de las habitaciones de los niños y encontraron que la mayor diversidad de microorganismos estaba inversamente relacionada con el asma, con independencia de si habitaban o no en una granja (14). En conjunto, estas investigaciones sugieren que la exposición a una amplia variedad de microorganismos puede proteger contra el asma, una hipótesis causal que debe ser probada en un estudio más sólido.

Preguntas de revisión

Lea las siguientes afirmaciones y marque la mejor respuesta.

2-1. Los registros de cáncer aportan 40 casos nuevos de cáncer de vejiga por cada 100 000 hombres por año. Los casos se obtuvieron de un recuento completo de todos los pacientes que desarrollaron cáncer de vejiga en varias regiones de EE.UU. y el número de hombres en riesgo se estimó de los datos del censo en esas regiones. Este es un ejemplo de qué tasa.

A. Prevalencia puntual.
B. Prevalencia de período o lápsica.
C. Densidad de incidencia.
D. Incidencia acumulada.
E. Tasa de complicaciones.

booksmedicos.org

2-2. El 60 % de los adultos de la población estadounidense tienen un colesterol sérico > 200 mg/dl (5,2 mmol/l). Este es un ejemplo de qué tasa.

A. Prevalencia puntual.
B. Tasa de complicaciones.
C. Densidad de incidencia.
D. Incidencia acumulada.
E. Prevalencia de período o lápsica.

2-3. Usted está leyendo un estudio sobre la prevalencia de las infecciones del cuello uterino y desea decidir si el estudio es científicamente sólido. ¿Cuál de las siguientes afirmaciones no es importante?

A. Se siguió a los participantes durante un período suficiente para que presentaran anemia.
B. El estudio se realizó en una muestra representativa de la población.
C. Todos los miembros de la población son mujeres.
D. Se define claramente el diagnóstico de infección cervical.
E. El estudio se realizó en una muestra de una población definida.

2-4. Una muestra probabilística en una población definida:

A. Se invalida por la sobrerrepresentación.
B. Es inferior a una muestra aleatoria.
C. No es representativa de la población.
D. Es una muestra representativa de la población sólo si hay suficientes personas en la muestra.

2-5. La incidencia de artritis reumatoide es aproximadamente de 40/100 000/año y la prevalencia es de alrededor de 1/100 personas. En promedio, ¿cuántos años persiste la enfermedad?

A. 10.
B. 25.
C. 35.
D. 40.
E. 50.

booksmedicos.org

2-6. ¿Cuál de los siguientes no es un estudio de cohorte?

A. La proporción de pacientes con cáncer gástrico que sobrevive 5 años.
B. El riesgo de los niños de desarrollar diabetes mellitus de acuerdo con su peso.
C. Las complicaciones de la vacunación contra la influenza entre los niños vacunados en 2011.
D. El curso temprano de la enfermedad en un grupo de pacientes que ahora están bajo el cuidado de un clínico.
E. Los pacientes ingresados en una unidad de cuidados intensivos con seguimiento para determinar si siguen vivos en el momento del alta hospitalaria.

2-7. Se obtiene una muestra para un estudio de incidencia sobre los errores en la medicación reclutando a cada décimo paciente que ingresa en el hospital. ¿Qué tipo de muestreo es este?

A. Muestreo estratificado.
B. Muestreo probabilístico.
C. Muestreo por conveniencia.
D. Muestreo aleatorio.
E. Sobrerrepresentación.

2-8. Los estudios de cohorte en niños con una primera convulsión febril han mostrado que tienen una probabilidad de uno entre tres de tener otra convulsión durante la infancia. ¿Qué tipo de tasa es esta?

A. Prevalencia puntual.
B. Tasa de complicaciones.
C. Incidencia acumulada.
D. Prevalencia de período o lápsica.
E. Densidad de incidencia.

2-9. ¿Cuál de los siguientes no incrementaría la incidencia observada de la enfermedad?

A. Esfuerzos más agresivos para detectar la enfermedad.
B. Un incremento verdadero en la incidencia.

booksmedicos.org

C. Una forma más sensible de detectar la enfermedad.
D. Una disminución del umbral para diagnosticar la enfermedad.
E. Estudiar una muestra más grande de la población.

2-10. La infección por un hongo, la coccidioidomicosis, es habitual en los desiertos del sureste de EE.UU. y México, pero infrecuente en otros sitios. ¿Cuál de los siguientes describe mejor esta infección?

A. Endémica.
B. Pandémica.
C. Incidental.
D. Epidémica.
E. Prevalente.

2-11. El 26 % de los adultos explicaron haber experimentado dolor lumbar que dura al menos 1 día en los últimos 3 meses. ¿Cuál de los siguientes describe mejor esta tasa?

A. Incidencia acumulada.
B. Densidad de incidencia.
C. Prevalencia puntual.
D. Tasa de complicaciones.
E. Prevalencia periódica.

2-12. ¿Cuál de los siguientes describe mejor una «población dinámica»?

A. Aumenta rápidamente su tamaño.
B. Es especialmente adecuada para los estudios de cohorte.
C. Las personas continuamente entran y abandonan la población.
D. Es la base para medir la incidencia acumulada.
E. Es el mejor tipo de población para una muestra aleatoria.

2-13. Para un estudio sobre la incidencia de escoliosis idiopática (una deformidad en la columna que se vuelve aparente después del nacimiento, con mayor frecuencia en la adolescencia), ¿cuál sería una cohorte apropiada?

A. Los niños nacidos en Carolina del Norte en 2012 y examinados hasta la vida adulta para detectar escoliosis.

booksmedicos.org

B. Los niños referidos a un cirujano ortopédico para recibir tratamiento.

C. Los niños de Carolina del Norte a quienes se les detectó escoliosis en una encuesta infantil.

D. Los niños que tienen escoliosis y están disponibles para el estudio.

E. Una muestra aleatoria de niños de una población de Carolina del Norte en la primavera de 2012.

2-14. ¿Para qué son especialmente útiles los estudios de prevalencia?

A. Describir la incidencia de la enfermedad.

B. Estudiar las enfermedades que se resuelven rápidamente.

C. Estimar la duración de la enfermedad.

D. Describir la proporción de personas en una población definida con la condición de interés.

E. Establecer causa y efecto.

2-15. El año pasado, 800 000 norteamericanos murieron por una enfermedad cardíaca o accidente cerebrovascular. ¿Cuál de las siguientes describe mejor esta estadística?

A. Densidad de incidencia.

B. Prevalencia puntual.

C. Incidencia acumulada.

D. Prevalencia periódica.

E. Ninguna de las anteriores.

Encontrará las respuestas en el apéndice A.

BIBLIOGRAFÍA

1. Roberts DE, Gupta G. Letter to the editor. New Engl J Med 1987;316:550.
2. Bryant GD, Norman GR. Expressions of probability: words and numbers. N Engl J Med 1980;302:411.
3. Toogood JH. What do we mean by “usually”? Lancet 1980;1:1094
4. Loftus EV Jr. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. Gastroenterology 2004;126:1504–1517.
5. Loftus EV Jr, Silverstein MD, Sandborn WJ, et al. Ulcerative colitis in Olmstead County, Minnesota, 1940-1993: incidence, prevalence, and survival. Gut 2000;46:336–343.
6. Sands K, Vineyard G, Platt R. Surgical site infections occurring after hospital discharge. J Infect

booksmedicos.org

Dis 1996;173:963–970.
7. Andrade L, Caraveo-Anduaga JJ, Berglund P, et al. The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) surveys. Int J Methods Psychiatr Res 2003;12:3–21.
8. Yawn BP, Saddier P, Wollan PC, et al. A population-based study of the incidence and complication rates of herpes zoster before zoster vaccine introduction. Mayo Clin Proc 2007;82: 1341–1349.
9. Flegal KM, Carroll MD, Ogden CL, et al. Prevalence and trends in obesity among US adults, 1999-2008. JAMA 2010; 303:235–241.
10. Jenkins C, Costello J, Hodge L. Systematic review of prevalence of aspirin induced asthma and its implications for clinical practice. BMJ 2004;328:434–437.
11. Wolf AMD, Wender RC, Etzioni RB, et al. American Cancer Society Guideline for the early detection of prostate cancer: update 2010. CA Cancer J Clin 2010;60:70–98.
12. Pang X, Zhu Z, Xu F, et al. Evaluation of control measures implemented in the severe acute respiratory syndrome outbreak in Beijing, 2003. JAMA 2003;290:3215–3221.
13. Center MM, Jemal A, Smith RA, et al. Worldwide variations in colorectal cancer. CA Cancer J Clin 2009;59:366–378.
14. Ege MJ, Mayer M, Normand AC, et al. Exposure to environmental microorganisms in childhood asthma. N Engl J Med 2011; 364:701–709.

booksmedicos.org

Capítulo 3

Anormalidad

... el significado médico de «normal» se ha perdido en la gran baraja de la estadística.

— Alvan Feinstein
1977

PALABRAS CLAVE

Datos nominales
Datos dicotómicos
Datos ordinales
Datos de intervalo
Datos continuos
Datos discretos
Validez
Exactitud
Apartados
Constructos
Escalas
Validez del contenido
Validez de criterio
Validez del constructo
Fiabilidad
Reproducibilidad
Precisión
Amplitud
Sensibilidad
Interpretabilidad

booksmedicos.org

Fracción de muestreo
Distribución de frecuencia
Tendencia central
Dispersión
Distribución desviada
Distribución normal
Regresión a la media

Los médicos dedican una gran parte del tiempo a distinguir entre lo que es «normal» y lo que es «anormal». El tamaño de la glándula tiroides, ¿es normal o está ligeramente aumentado? El soplo cardíaco, ¿es inocente (sin importancia para la salud) o es un signo de valvulopatía? Una concentración sérica de fosfatasa alcalina ligeramente elevada ¿es un indicio de hepatopatía, de enfermedad de Paget asintomática, o es un hecho sin importancia?

Cuando los médicos se enfrentan a un caso claramente diferente de lo habitual no tienen dificultades para distinguirlo de los demás. Todos hemos visto imágenes en los libros de texto de diagnósticos clínicos que muestran grandes hepatoesplenomegalias, enormes bocios o manos gravemente deformadas por la artritis reumatoide. No es necesaria una gran habilidad para reconocer este grado de anormalidad, pero los médicos rara vez se enfrentan a esta situación. Lo más frecuente es que los médicos tengan que realizar distinciones más sutiles entre lo normal y lo anormal. Es entonces cuando adquieren importancia la habilidad y una base conceptual para decidir.

Las decisiones sobre lo que es anormal son especialmente difíciles entre los pacientes relativamente no seleccionados, que en general suelen encontrarse fuera de los hospitales. Cuando los pacientes ya han sido seleccionados para recibir una atención especial, como ocurre en la mayoría de los centros de referencia, habitualmente suele estar claro que algo va mal. En tal caso, las tareas consisten en precisar el diagnóstico y tratar el problema. No obstante, en los ámbitos de atención primaria y en los servicios de urgencias, los pacientes con manifestaciones patológicas sutiles se mezclan con los que refieren afecciones cotidianas de la gente que está sana. Es imposible hacer un seguimiento activo de todas estas afecciones. De los muchos pacientes con dolor abdominal, ¿quiénes sufren una gastroenteritis que se resolverá de forma espontánea, y quiénes una apendicitis inicial? De los pacientes con dolor de garganta y ronquera, ¿quiénes sufren una faringitis vírica y quiénes una rara, pero potencialmente mortal, epiglotitis por

booksmedicos.org

Haemophilus influenzae? Estos son ejemplos de lo difícil e importante que puede llegar a ser la distinción entre los diferentes tipos de anormalidades.

El objetivo de distinguir entre lo normal y lo anormal es separar las observaciones clínicas que deben constituir la base para la actuación de las que sencillamente pueden quedar registradas. Las observaciones que se consideran normales suelen describirse como «dentro de los límites normales», «insignificantes» o «sin contribución», y quedan enterradas en la historia clínica del paciente. Los hallazgos anormales se incluyen en una lista de problemas o bajo títulos como «Impresiones» o «Diagnósticos», y conforman la base para la actuación.

Tabla 3-1

Resumen de datos clínicos: lista de problemas de un paciente y datos en los que estos se basan

Problemas	Datos brutos
Infarto agudo de miocardio	Dolor torácico, troponina 40 µg/l (> percentil 99 en el límite superior de referencia), nueva elevación del segmento ST en las derivaciones II, III y AVF
Hipertensión	Diferentes mediciones de la presión arterial (mm Hg): 145/92, 149/93, 142/91
Diabetes mellitus	Varias mediciones de glucosa plasmática en ayunas (mg/dl): 138, 135, 129
Insuficiencia renal	Creatinina sérica 2,7 mg/dl
Enfermedad pulmonar obstructiva	Volumen espiratorio forzado en 1 s (FEV1)/capacidad vital forzada (CVF) < 0,70

Calificar los hallazgos clínicos sencillamente como normales o anormales denota una clara falta de precisión y da lugar a clasificaciones erróneas. El motivo de estos enfoques es que a menudo no resulta práctico o es innecesario considerar los datos brutos con todo detalle. Como señaló Bertrand Russell: «Para ser perfectamente inteligible, uno tiene que ser al menos algo inexacto, y para ser perfectamente exacto, uno tiene que ser ininteligible». A menudo los médicos prefieren inclinarse por la inteligibilidad (para ellos mismos y para los demás médicos), incluso a expensas de cierta exactitud. Otra razón para simplificar los datos es que cada

booksmedicos.org

aspecto del cometido de un médico termina en una decisión: seguir con la evaluación o esperar, empezar el tratamiento o tranquilizar al paciente. En estas circunstancias es necesario algún tipo de clasificación como «presente» o «ausente».

La tabla 3-1 es un ejemplo de cómo las expresiones relativamente sencillas de anormalidad proceden de datos clínicos más complejos. A la izquierda hay una lista de problemas habituales, un enunciado de los problemas médicos importantes para el paciente. A la derecha se exponen algunos de los datos utilizados como base para tomar la decisión de etiquetarlos como problemas. Las conclusiones de los datos, representadas por la lista de problemas, son por supuesto objetables. Por ejemplo, la media de las cuatro determinaciones de la presión arterial diastólica es de 92 mm Hg. Puede que algunos afirmen que tal nivel de presión arterial no justifica la calificación de «hipertensión arterial», ya que no se trata de una cifra especialmente elevada, y el hecho de informar a los pacientes de que están enfermos, así como de recetar fármacos, presenta algunos inconvenientes. Puede que otros consideren que la calificación es adecuada, si se tiene en cuenta que este nivel de presión arterial se asocia a un aumento del riesgo de enfermedad cardiovascular, y que dicho riesgo puede reducirse mediante tratamiento y esta calificación es consistente con las pautas de tratamiento. Aunque poco precisa, la lista de problemas sirve como base para tomar decisiones (sobre el diagnóstico, el pronóstico y el tratamiento). Y deben tomarse decisiones clínicas, ya sea de forma activa (mediante pruebas diagnósticas adicionales y tratamiento) o de forma pasiva (sin intervención).

Este capítulo describe algunos de los métodos que utilizan los médicos para distinguir entre lo normal y lo anormal. Primero, será necesario considerar cómo se miden, cómo varían y cómo se resumen los fenómenos biológicos. Posteriormente, se podrá considerar cómo se utilizan estos datos como base de los juicios de valor que permiten determinar qué merece la pena denominar anómalo.

TIPOS DE DATOS

Las mediciones de los fenómenos clínicos dan lugar a tres tipos de datos: nominales, ordinales y de intervalo.

Datos nominales

booksmedicos.org

Los **datos nominales** se producen en categorías sin ningún orden natural. Las características determinadas por un pequeño grupo de genes (como el grupo sanguíneo ABO o el sexo) o que constituyen episodios aislados, pero espectaculares (como la muerte, la diálisis o una intervención quirúrgica), son ejemplos de datos nominales. Dichos datos pueden incluirse en diferentes categorías sin preocuparse demasiado por clasificarlos de forma errónea. Los datos nominales que se dividen en dos categorías (como presente/ausente, sí/no, vivo/ muerto) se denominan **dicotómicos.**

Datos ordinales

Los **datos ordinales** poseen cierto orden natural, como por ejemplo todo lo situado entre pequeño y grande, o entre bueno y malo, si bien el tamaño de los intervalos entre categorías no tiene por qué ser definido o detallado. Algunos ejemplos clínicos son el edema de las extremidades inferiores entre 1+ y 4+, los soplos cardíacos entre los grados I (audible sólo tras un esfuerzo especial) y VI (audible sin necesidad de estetoscopio), y la fuerza muscular entre los grados 0 (sin movimiento) a 5 (fuerza normal). Algunas escalas ordinales son complejas. La Food and Drug Administration (FDA) ha clasificado el riesgo de sufrir anomalías congénitas a causa de la utilización de fármacos durante el embarazo en una escala que consta de cinco categorías que van desde la A, «sin efectos adversos en los humanos»; B, un efecto adverso en estudios en animales no confirmado en estudios controlados en mujeres, o bien «sin efecto en animales sin información sobre el efecto en humanos»; C, «efectos adversos en animales sin información sobre el efecto en humanos o sin información disponible de animales o humanos», y D, «efectos adversos en humanos o probables efectos en humanos debido a los efectos adversos en animales»; hasta la X, «efectos adversos en humanos o animales con uso contraindicado durante el embarazo» (1).

Datos de intervalo

Para los **datos de intervalo,** existe un orden inherente y el intervalo entre valores sucesivos es idéntico, sin importar su situación dentro de la escala. Existen dos tipos de datos de intervalo. Los **datos continuos** pueden adquirir cualquier valor dentro de un intervalo continuo, sin importar la forma en que se notifiquen. La mayor parte de la bioquímica sérica, el peso, la presión arterial y la presión parcial del oxígeno en sangre arterial son algunos

booksmedicos.org

ejemplos de datos continuos. En la práctica, la medición y la descripción de las variables continuas pueden limitarse a un número determinado de puntos dentro del intervalo continuo, a menudo representados por números enteros, ya que la precisión de la medición o su utilización no justifican un mayor detalle. Por ejemplo, la lectura de una medición concreta de la glucemia puede ser de 193,2846573... mg/ml, pero se documenta sencillamente como 193 mg/ml. Los **datos discretos** sólo pueden adquirir valores determinados, y se expresan como recuentos. El número de embarazos y partos de una mujer, así como el número de crisis de migraña que sufre un paciente al mes, son ejemplos de datos discretos.

Precisamente, es para los datos ordinales y de intervalo cuando surge la pregunta: «¿Dónde termina lo normal y dónde empieza lo anormal?» Por ejemplo, ¿en qué momento puede afirmarse que una próstata normal de tamaño grande ha crecido lo bastante como para dejar de considerarse normal? Los médicos son libres de elegir cualquier punto de corte. Algunas de las razones para estas elecciones se tratarán más adelante en este capítulo.

RENDIMIENTO DE LAS MEDICIONES

Con independencia del tipo de medición, su rendimiento puede describirse de varias formas.

Validez

La **validez** es el grado en que los datos miden aquello que pretendían medir, es decir, el grado en que los resultados de la medición se corresponden con el estado real del fenómeno medido. Un sinónimo es **exactitud.**

Es relativamente fácil comprobar la validez de las observaciones clínicas que pueden medirse por medios físicos. La medición observada se compara con alguna referencia aceptada. Por ejemplo, el sodio sérico puede medirse mediante un instrumento calibrado recientemente con unas soluciones con concentraciones conocidas de sodio. Las mediciones analíticas suelen someterse a amplias y repetidas comprobaciones para determinar su validez. Por ejemplo, es una práctica habitual controlar la exactitud de las mediciones de la glucemia comparando las lecturas con referencias elevadas y bajas al principio de cada día, antes de que cada técnico empiece su jornada, y después de cualquier cambio en los procedimientos, como un nuevo frasco de reactivo o una nueva pila para el instrumento. De igual modo, puede medirse

booksmedicos.org

la exactitud de una gammagrafía pulmonar realizada para detectar un émbolo pulmonar, comparándola con una angiografía pulmonar en la que se visualiza directamente la anatomía de las arterias pulmonares. La validez de un hallazgo durante una exploración física puede comprobarse al compararlo con los resultados de la cirugía o de las exploraciones radiológicas.

Tabla 3-2

Cuestionario CAGE para la detección del consumo excesivo o abusivo de alcohol[a]
¿Ha sentido alguna vez la necesidad de dejar *(Cut)* de beber?
¿Le ha molestado *(Annoyed)* que la gente haya criticado su forma de beber?
¿Se ha sentido culpable *(Guilty)* por su forma de beber?
¿Alguna vez ha sentido que necesitaba un primer trago por la mañana *(Eye opener)* para calmar sus nervios o para librarse de la resaca?
Una respuesta «afirmativa» sugiere la necesidad de una valoración más detenida
Un mínimo de dos respuestas «afirmativas» expresan una estrecha relación con abuso, la dependencia o con ambos

[a]Otras pruebas, como AUDIT, son útiles para detectar patrones de consumo menos graves que puedan responder a una simple asesoría.
Adaptado de Ewing JA. Detecting alcoholism: the CAGE questionnaire. JAMA 1984;252:1905-1907.

Algunas otras mediciones clínicas, como el dolor, las náuseas, la disnea, la depresión y el miedo, no pueden comprobarse de forma física. En las consultas, la información relacionada con estos fenómenos suele obtenerse de modo informal a través de la anamnesis, aunque existen otros métodos más formales y estandarizados que se utilizan en la investigación, como las entrevistas estructuradas y los cuestionarios. En este caso, se diseñan preguntas concretas **(apartados)** para medir fenómenos determinados (como síntomas, sentimientos, actitudes, conocimientos, creencias) denominados **constructos,** y dichos apartados se agrupan en **escalas.** La tabla 3-2 muestra una de estas escalas; se trata de un breve cuestionario utilizado para detectar el abuso y dependencia del alcohol.

Se utilizan tres estrategias generales para comprobar la validez de las mediciones que no pueden verificarse directamente.

booksmedicos.org

Validez del contenido

La **validez del contenido** es el grado en que un método concreto de medición incluye todas las dimensiones del constructo que se pretende medir y nada más. Por ejemplo, una escala para medir el dolor tendrá validez de contenido si incluye preguntas sobre dolor sordo, dolor pulsátil, presión, dolor urente y escozor, pero no sobre prurito, náuseas y hormigueo.

Validez de criterio

Existe **validez de criterio** cuando las mediciones predicen un fenómeno que se puede observar directamente. Por ejemplo, es posible verificar si las respuestas en una escala que mide el dolor guardan una relación predecible con el dolor de una intensidad conocida: dolor leve producido por una abrasión sin importancia, dolor moderado causado por una cefalea común y una úlcera péptica, y dolor intenso debido a un cólico nefrítico. También es posible observar que las respuestas a una escala que mide el dolor están relacionadas con otras manifestaciones observables vinculadas a la intensidad del dolor, como sudoración, gemidos y el hecho de solicitar fármacos contra el dolor.

Validez del constructo

Existe **validez del constructo** cuando la medición está relacionada de forma coherente con otras mediciones, que, al igual que esta, no pueden comprobarse físicamente, y que se consideran parte del mismo fenómeno. Así, por ejemplo, uno puede tener más confianza en la validez de constructo de una escala que mide la depresión cuando esta esté relacionada con la fatiga y la cefalea (otros constructos que se consideran diferentes pero relacionados con la depresión).

La validez de una escala no es algo que esté, como se afirma a menudo, presente o ausente. Por el contrario, con estas estrategias, es posible adoptar una posición o construir un escenario a favor o en contra de la validez de una escala en las condiciones en las que se utiliza, de forma que se pueda convencer a los demás de que la escala es más o menos válida.

Debido a su selección y formación, los médicos tienden a preferir el tipo de mediciones precisas que permiten las ciencias físicas y biológicas, y pueden evitar o descartar las otras, especialmente para la investigación. A pesar de todo, el alivio de los síntomas y el fomento de la satisfacción y de la sensación de bienestar figuran entre los resultados más importantes de la

booksmedicos.org

asistencia a los pacientes, y constituyen las preocupaciones centrales tanto de estos como de los médicos. La investigación también debe incluir dichas preocupaciones, a fin de guiar las decisiones clínicas y evitar que la propia investigación distorsione la imagen que ella misma ha dibujado de la medicina.

Tal como Feinstein afirmó (2):

> El término «duro» suele aplicarse a los datos que son fiables y preferiblemente dimensionales (p. ej., datos analíticos, datos demográficos y costos económicos). Pero el rendimiento clínico, la conveniencia, la anticipación y los datos familiares son «suaves». Dependen de afirmaciones subjetivas, habitualmente expresadas con palabras y no con números, realizadas por las personas que actúan como observadoras y como observadas.
>
> Para evitar este tipo de datos suaves, los resultados del tratamiento suelen restringirse a la información analítica que pueda ser objetiva, dimensional y fiable, aunque también quede deshumanizada. Si nos dicen que el colesterol sérico es de 230 mg/ml, que la radiografía torácica muestra una cardiomegalia y que el electrocardiograma presenta ondas Q, no sabremos si el objeto de que se trata es un perro o una persona. Si nos dijeran que se ha recuperado la capacidad para trabajar, que el medicamento tiene un sabor agradable y es fácil de tomar, y que la familia se muestra satisfecha con los resultados, reconoceríamos un grupo de respuestas humanizadas.

Fiabilidad

La **fiabilidad** es el grado en que mediciones repetidas de un fenómeno estable, realizadas por personas e instrumentos diferentes, en momentos y lugares distintos, expresan resultados similares. **Reproducibilidad** y **precisión** son otros términos que designan la misma propiedad.

La fiabilidad de las determinaciones analíticas se establece mediante mediciones repetidas (p. ej., del mismo suero o muestra tisular) llevadas a cabo, a veces, por personas diferentes y con instrumentos distintos. La fiabilidad de los síntomas puede comprobarse demostrando que diferentes observadores en condiciones diferentes los describen de forma similar.

En la figura 3-1 se muestran de forma simplificada las relaciones entre la fiabilidad y la validez. Las mediciones pueden ser exactas (válidas) y fiables (precisas), tal como muestra la figura 3-1 A. Las mediciones pueden ser muy fiables, pero inexactas si están sistemáticamente fuera del objetivo o valor a medir de la puntuación deseada, como sucede en la figura 3-1 B. Por otro lado, las mediciones pueden ser válidas de promedio, pero no fiables, puesto que se encuentran muy dispersas en torno al valor real, tal como muestra la figura 3-1 C. Finalmente, las mediciones pueden ser inválidas e imprecisas, como se muestra en la figura 3-1 D. Un número reducido de mediciones con

booksmedicos.org

una fiabilidad deficiente también tendrán poca validez, ya que es muy posible que se alejen del valor real sólo por efecto del azar. Por lo tanto, la fiabilidad y la validez no son conceptos del todo independientes. En general, una medición poco fiable no puede ser válida, y una medición válida debe ser fiable.

Amplitud

Puede que un instrumento no registre los valores muy elevados o muy bajos del fenómeno que se está midiendo, es decir, tiene una **amplitud** limitada, y la información que proporciona también lo es. Por ejemplo, la Escala de actividades básicas de la vida diaria que mide la capacidad de los pacientes para vestirse, comer, caminar, utilizar el retrete, mantener la higiene y trasladarse de una cama o silla no mide la capacidad de leer, escribir o tocar el piano (actividades que pueden ser muy importantes para determinados pacientes).

Sensibilidad

Un instrumento demuestra su **sensibilidad** en la medida en que, al cambiar las condiciones, cambian los resultados. Por ejemplo, la Escala de la New York Heart Association, con clases de I a IV (sin síntomas de insuficiencia cardíaca ni limitaciones para realizar actividad física ordinaria, síntomas leves con ligera limitación para realizar actividad física ordinaria, limitación importante para realizar actividad física ordinaria a causa de fatiga, palpitaciones o disnea o incapacidad para realizar cualquier actividad física, incluso en reposo, a causa de los síntomas) no es sensible a los cambios sutiles que se producen en una insuficiencia cardíaca congestiva, precisamente aquellos que pueden ser importantes para los pacientes. Sin embargo, las mediciones de la fracción de eyección realizadas mediante ecocardiografía pueden detectar cambios tan sutiles que no los perciben ni los pacientes.

booksmedicos.org

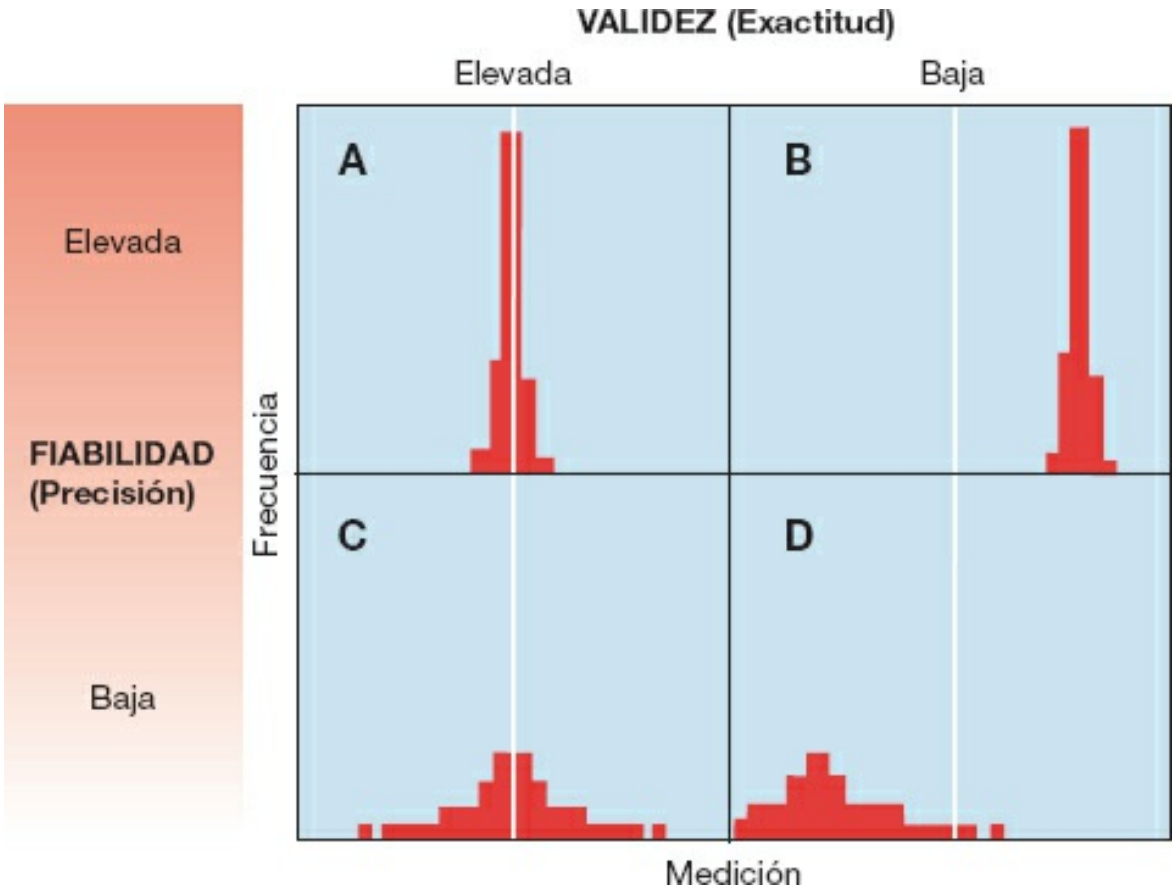


Figura 3-1 ■ Validez y fiabilidad. A. Validez elevada y fiabilidad elevada. **B.** Validez baja y fiabilidad elevada. **C.** Validez elevada y fiabilidad baja. **D.** Validez baja y fiabilidad baja. Las *líneas de puntos* son los valores reales.

Interpretabilidad

Los médicos aprenden a interpretar el significado de una PCO_2 de 50 mm Hg o de una glucemia de 460 mg/ml mediante la experiencia, tras contrastar repetidamente las condiciones de los pacientes y su evolución clínica con los resultados de tales pruebas. Sin embargo, las escalas basadas en cuestionarios pueden tener poco significado intuitivo para los médicos y los pacientes que no las utilizan de forma regular. Para superar este inconveniente relacionado con la **interpretabilidad,** los investigadores pueden «anclar» una correspondencia entre los valores de la escala y las situaciones conocidas. Para ayudar a los médicos a interpretar los valores de la escala, los números están ligados a descripciones de uso diario. Por ejemplo, los valores de la Escala de Karnofsky *(Karnofsky Performance Status Scale)*, utilizada habitualmente en los estudios para medir la capacidad funcional de los pacientes con cáncer tratados con quimioterapia, oscilan ente 100 (normal) y

booksmedicos.org

0 (muerte). Así, ¿cuál es la gravedad al obtener un valor de 60? Con un valor de 60 en la escala, los pacientes precisan ayuda ocasional, pero son capaces de solventar por sí mismos la mayor parte de sus necesidades personales.

VARIACIÓN

La variación global es la suma de la variación relacionada con el propio acto de la medición, de las diferencias biológicas *en el interior* de las personas de un momento a otro y de las diferencias biológicas entre las personas (tabla 3-3).

Tabla 3-3

Fuentes de variación	
Fuente de variación	**Definición**
Variación de la medida	
Instrumento	Medios utilizados para realizar la medición
Observador	Persona que lleva a cabo la medición
Variación biológica	
Dentro de cada persona	Cambios que se producen en una persona con el tiempo y la situación
Entre personas	Diferencias biológicas entre una persona y otra

Variación derivada de la medición

Todas las observaciones están sujetas a variaciones, debido al funcionamiento de los instrumentos y a los observadores que intervienen en las mediciones. Las condiciones en que se realiza la medición pueden dar lugar a un resultado sesgado (falta de validez) o, sencillamente, a un error aleatorio (falta de fiabilidad). Esta fuente de variación puede reducirse si las mediciones se realizan con sumo cuidado y si se siguen protocolos estandarizados. Sin embargo, cuando las mediciones dependen del juicio humano, y no de las máquinas, la variación que se genera puede ser especialmente amplia y difícil de controlar.

booksmedicos.org

Ejemplo

Los hallazgos en la radiografía de tórax se utilizan como parte del diagnóstico de la lesión pulmonar aguda y el síndrome de dificultad respiratoria aguda (LPA-SDRA), una grave enfermedad pulmonar con hipoxemia arterial que a menudo requiere intubación. Pero, ¿los especialistas en estos trastornos respiratorios leen las radiografías de manera similar? En un estudio, 21 expertos de atención crítica pulmonar examinaron 28 radiografías de tórax de pacientes gravemente enfermos con hipoxemia y decidieron si las radiografías cumplían los criterios radiográficos para el diagnóstico de LPA-SDRA. El porcentaje de radiografías consideradas positivas para el diagnóstico varió de 36-71% entre los expertos (fig. 3-2), con una diferencia entre los clínicos de más de dos veces entre los porcentajes positivo menor y mayor. Las radiografías con el mayor grado de acuerdo entre los expertos mostraron anomalías en todos los cuadrantes pulmonares, mientras que aquellas con anomalías limitadas al cuadrante pulmonar inferior representaron el mayor grado de variabilidad (3).

También pueden producirse variaciones en las mediciones porque se realizan en una sola muestra del fenómeno que se describe, lo que puede conllevar una representación errónea del conjunto. A menudo, la **fracción de muestreo** (fracción del conjunto incluida en la muestra) es muy pequeña. Por ejemplo, una biopsia hepática sólo representa un 1/100 000 del hígado. Puesto que se examina una parte tan minúscula del conjunto, es posible que se produzca una variación considerable entre una muestra y otra.

Si las mediciones se realizan con métodos diferentes, como laboratorios, técnicos o instrumentos distintos, puede que algunas sean poco fiables o proporcionen resultados sistemáticamente diferentes del valor correcto, lo que contribuiría a la dispersión de los valores obtenidos.

booksmedicos.org

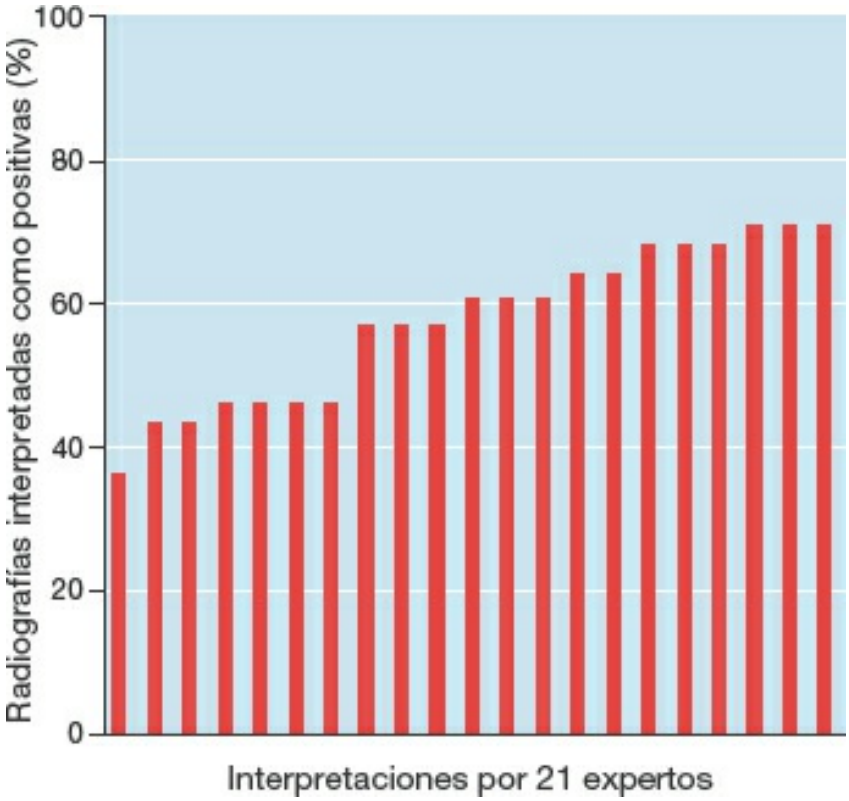


Figura 3-2 ■ Variabilidad del observador. Variabilidad entre los 21 especialistas que interpretan las radiografías de tórax para descartar lesión pulmonar aguda y síndrome de dificultad respiratoria aguda. El porcentaje de radiografías consideradas positivas para el diagnóstico varió de 36-71% entre los expertos. (De Rubenfeld GD, Caldwell E, Granton J, et al. Interobserver variability in applying a radiographic definition for ARDS. Chest 1999;116:1347-1353.)

Variación derivada de las diferencias biológicas

Los cambios biológicos en el interior de cada persona a lo largo del tiempo también pueden dar lugar a variaciones. La mayor parte de los fenómenos biológicos cambian de un momento a otro. Una medición en un punto del tiempo puede que no represente el valor habitual de dichas mediciones.

Ejemplo

Los médicos calculan la frecuencia de los latidos ventriculares prematuros (LVP) para poder determinar la necesidad de un tratamiento y la eficacia de este con más facilidad. Por razones prácticas, pueden llevar a cabo este procedimiento mediante observaciones relativamente breves, ya sea tomando el pulso del paciente durante 1 min o revisando el registro de un electrocardiograma durante algunos segundos. Sin embargo, la frecuencia de los LVP en un paciente concreto varía a lo largo del tiempo. Para obtener una muestra mayor con el fin de estimar la frecuencia de LVP, se desarrolló un monitor portátil que rastrea las despolarizaciones

booksmedicos.org

ventriculares prematuras (DVP) en un electrocardiograma. Los primeros estudios encontraron que la monitorización, incluso durante largos períodos, puede inducir a error. La figura 3-3 muestra las observaciones realizadas en un paciente con DVP, similar a otros pacientes estudiados (4). Las DVP oscilaron entre menos de 20 a 380 por hora, en función del día y el momento del mismo, a lo largo de 3 días. Los autores concluyeron que «para distinguir una disminución de la frecuencia de los LVP atribuible sólo a la intervención terapéutica, y no a la variación biológica o espontánea, se requería una reducción superior al 83% en la frecuencia de LVP, si sólo se comparaban dos períodos de monitorización de 24 h». Unos períodos de observación más cortos podrían ser más propensos al error debido a la variación biológica. Para superar esta variación biológica, ahora los dispositivos modernos son capaces de monitorizar el ritmo cardíaco durante períodos más prolongados.

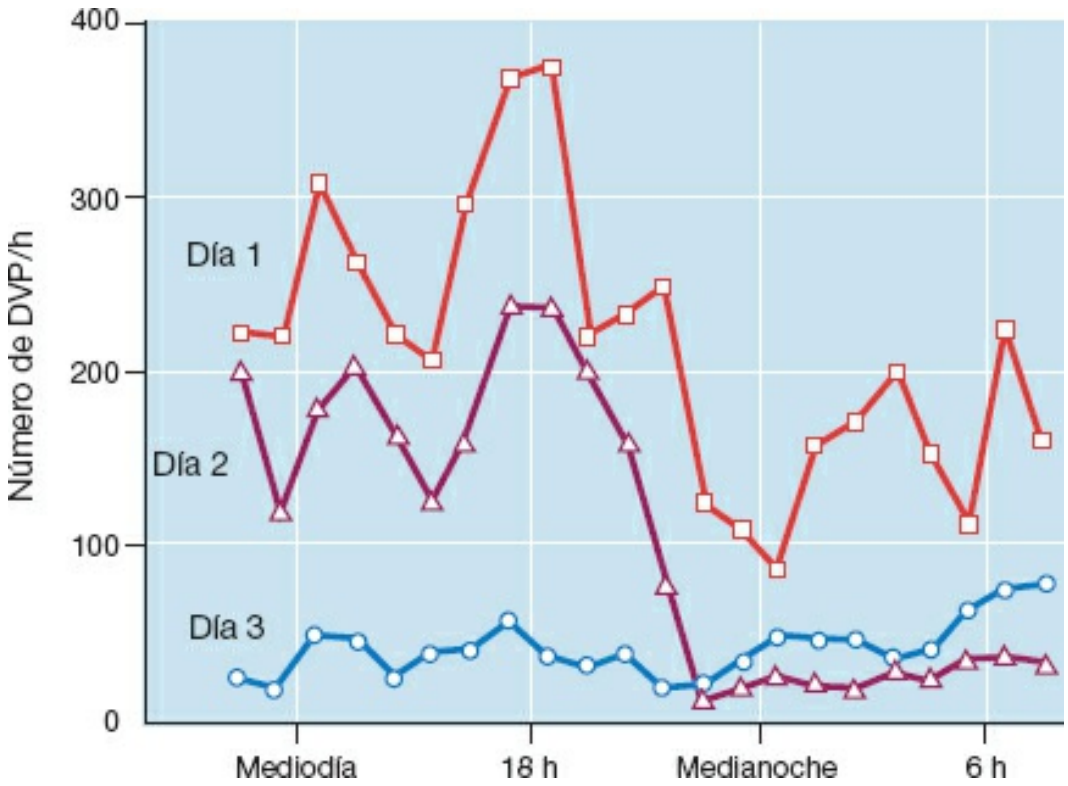


Figura 3-3 ■ Variabilidad biológica. Número de despolarizaciones ventriculares prematuras (DVP) por hora en un paciente sin tratamiento previo durante 3 días consecutivos. (De Morganroth J, Michelson EL, Horowitz LN, et al. Limitations of routine long-term electrocardiographic monitoring to assess ventricular ectopic frequency. Circulation 1978;58:408-414.)

Variación total

Las diferentes fuentes de variación son acumulativas. La figura 3-4 ilustra este hecho en la medición de la presión arterial. Cuando se observa la distribución de una población, la variación de la medición en pacientes concretos se suma a la variación que se produce en dichos pacientes entre un momento y otro, que a su vez se suma a la variación entre pacientes

booksmedicos.org

diferentes. La variación de la medición contribuye relativamente poco, aunque equivale unos 12 mm Hg entre diferentes observadores. Sin embargo, la presión arterial de cada paciente varía en gran medida de un momento a otro del día, por lo que cualquier lectura de la misma puede no representar el valor habitual de ese paciente. Una gran parte de esta variación no es aleatoria: en general, la presión arterial es más elevada cuando las personas están despiertas o excitadas, cuando acuden al médico o cuando toman fármacos de libre dispensación para el resfriado. Obviamente, lo que más nos interesa es saber cómo se compara la presión arterial de un paciente con la del resto de los mismos, sobre todo si el valor obtenido está relacionado con complicaciones de la hipertensión arterial y con la eficacia del tratamiento.

A pesar de todas las fuentes de variación que pueden distorsionar la medición de la presión arterial, las diferencias biológicas entre los individuos son la causa principal de la variación en el caso de la presión arterial, tanto es así que varios estudios han encontrado que incluso una sola medición casual de la presión arterial puede predecir la enfermedad cardiovascular subsecuente entre la población.

Efectos de la variación

Otro modo de concebir la variación es desde la perspectiva de su efecto neto sobre la validez y la fiabilidad de una medición, y de lo que se pueda hacer al respecto.

La variación aleatoria, causada, por ejemplo, por instrumentos inestables o por muchos observadores con diferentes sesgos que tienden a compensarse entre sí, no da lugar, finalmente, a ninguna representación errónea neta de la situación real de un fenómeno, aunque algunas mediciones induzcan a error. La inexactitud generada por la variación aleatoria puede reducirse tomando la media de una muestra mayor de lo que se está midiendo, por ejemplo, contando un mayor número de células en un frotis de sangre, examinando una mayor área de sedimento urinario o estudiando un mayor número de pacientes. Asimismo, el grado de variación aleatoria puede estimarse mediante métodos estadísticos (v. cap. 11).

booksmedicos.org

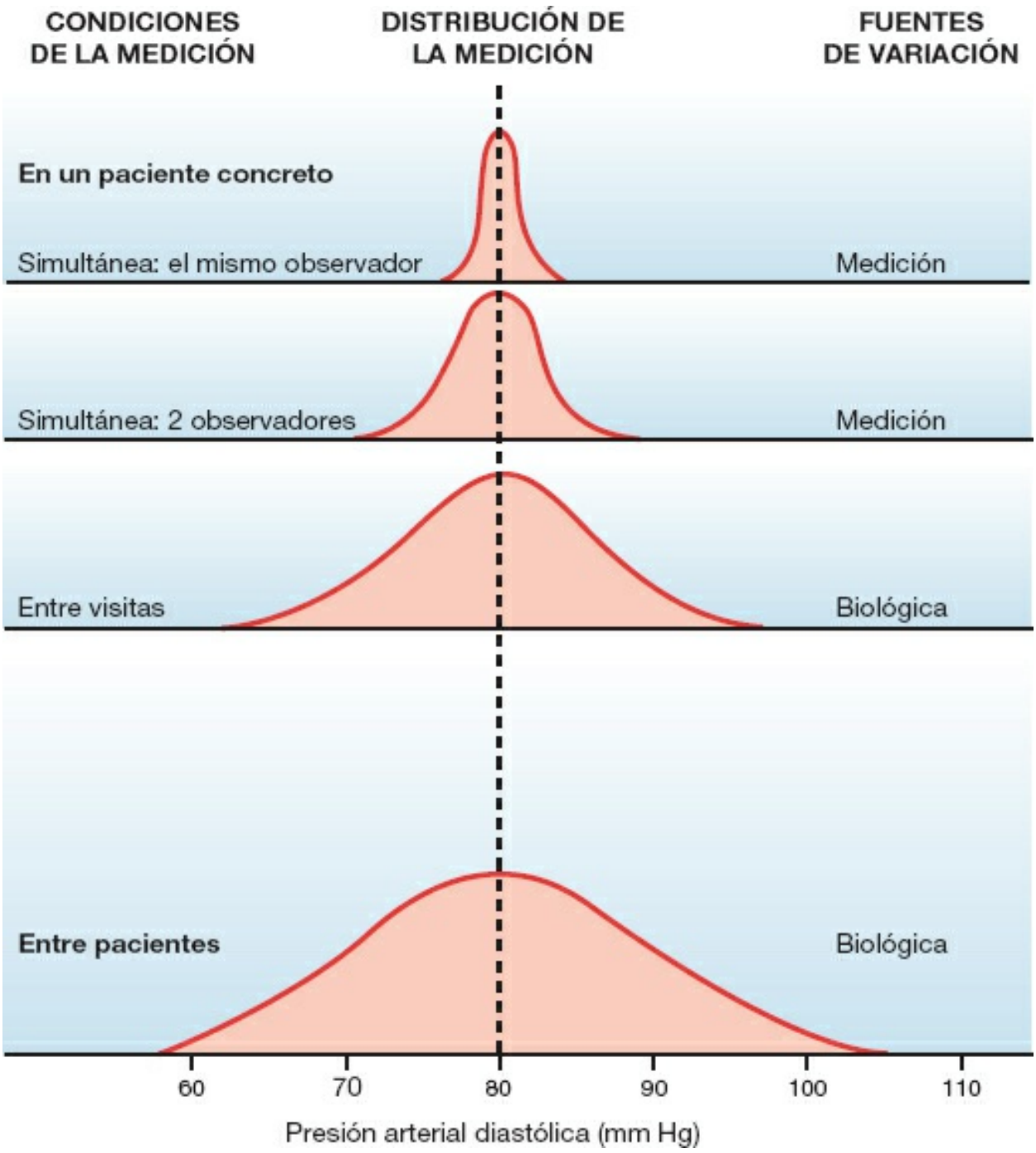


Figura 3-4 ■ Fuentes de variación en la medición de la presión arterial diastólica (fase V). La *línea de puntos* indica la presión arterial real. Existen múltiples fuentes de variación, incluyendo dentro y entre los pacientes, así como la variación intra e interobservador, que contribuyen a los resultados de la medición de la presión arterial.

Por otro lado, los resultados sesgados son sistemáticamente diferentes del valor real, con independencia del número de veces que se repitan. Por ejemplo, todos los valores elevados para las DVP descritas en la figura 3-3 se

booksmedicos.org

registraron el primer día, y la mayor parte de los valores bajos el tercero. Los días fueron estimaciones sesgadas unos respecto a otros debido a la variación en la frecuencia de las DVP entre un día y otro.

DISTRIBUCIONES

Los datos que se miden mediante escalas de intervalo se presentan a menudo como una figura, denominada **distribución de frecuencia,** en la que se muestra el número (o la proporción) de un grupo definido de personas que poseen los diferentes valores de aquello que se mide. La figura 3-5 muestra la distribución del recuento de neutrófilos en hombres y mujeres sanos de raza negra.

Descripción de las distribuciones

La presentación de los datos de intervalo como una distribución de frecuencia expresa la información con mucho detalle, pero a menudo es conveniente resumir las distribuciones. De hecho, esta acción es obligada cuando se presenta y se compara un gran número de distribuciones.

Para ello, se utilizan dos propiedades básicas de las mismas: la **tendencia central**, o punto central de la distribución, y la **dispersión,** o grado en que se encuentran repartidos los valores. En la figura 3-5 se ilustran varias formas de expresar la tendencia central y la dispersión, así como sus ventajas e inconvenientes. Estos conceptos también aparecen resumidos en la tabla 3-4.

booksmedicos.org

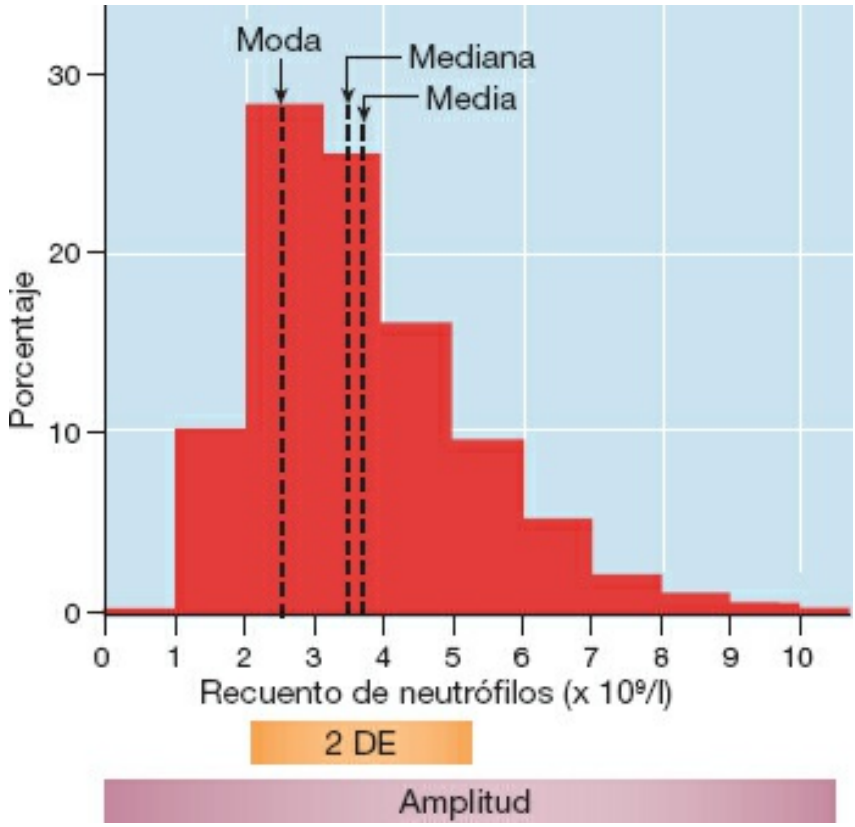


Figura 3-5 ■ Medidas de tendencia central y dispersión. Distribución de los niveles de PSA en varones supuestamente sanos. La distribución del recuento de neutrófilos en una muestra nacional de hombres de raza negra de 18 años o más. (Los autores encontraron que el recuento de neutrófilos fue menor y la neutropenia fue más habitual en los hombres de raza negra que en los caucásicos). (De Hsieh MM, Everhart JE, Byrd-Holt DD, et al. Prevalence of neutropenia in the U.S. population: Age, sex, smoking status and ethnic differences. Ann Intern Med 2007;146:486-492.)

Distribuciones reales

Las distribuciones de los fenómenos clínicos pueden adoptar muchas formas diferentes. En la figura 3-6 se muestran las distribuciones de frecuencia de cuatro tipos de análisis de sangre habituales (potasio, fosfatasa alcalina, glucosa y hemoglobina). En general, la mayor parte de los valores aparecen próximos al centro de la distribución y, con excepción de la parte central de las curvas, no se observan protuberancias ni irregularidades. Los extremos elevados y bajos de cada distribución se extienden formando colas, una de las cuales suele ser más alargada que la otra (es decir, las curvas están **desviadas** hacia el extremo más alargado). Algunas distribuciones están desviadas hacia los valores más elevados, mientras que otras lo están hacia los valores más bajos. En otras palabras, todas estas distribuciones son unimodales (poseen una sola protuberancia) y presentan forma de campana, aunque no

booksmedicos.org

necesariamente simétrica. Por lo demás, no se parecen entre sí.

La distribución de los valores de muchas pruebas de laboratorio cambia en función de las características de los pacientes, como la edad, el sexo, la raza y la alimentación. Por ejemplo, la distribución de una de estas pruebas, el nitrógeno ureico en sangre (BUN, prueba para medir la función renal), cambia con la edad. Una concentración de BUN de 25 mg/dl sería un valor anormalmente elevado para una persona joven, pero no representaría un valor destacable en una persona de 80 años.

Tabla 3-4

Expresiones de tendencia central y dispersión

Expresión	Definición	Ventajas	Inconvenientes
Tendencia central			
Media	$\frac{\text{Suma de los valores de las observaciones}}{\text{Número de observaciones}}$	Idónea para la manipulación matemática	Se ve afectada por los valores extremos
Mediana	Punto de la distribución en el que el número de observaciones situado por encima es el mismo que el situado por debajo	Los valores extremos no la afectan fácilmente	Poco adecuada para la manipulación matemática
Moda	El valor obtenido con más frecuencia	Sencillez de significado	En ocasiones no existen valores o bien hay muchos
Dispersión			
Amplitud	Valores situados entre el valor más bajo y el más alto de una distribución	Incluye todos los valores	Se ve enormemente afectada por los valores extremos
Desviación estándar[a]	Valor absoluto de la diferencia media entre cada valor respecto a la media[a]	Idónea para la manipulación matemática	En las distribuciones no gaussianas (v. a continuación) no describe una proporción conocida de las observaciones
Percentil, decil, cuartil, etc.	Proporción de todas las observaciones situadas entre valores específicos	Describe la «infrecuencia» de un valor No realiza suposiciones sobre la forma de la distribución	Poco adecuados para la manipulación estadística

[a] $\sqrt{\frac{\Sigma(X - \bar{X})^2}{N - 1}}$, donde X = cada observación; $\bar{X}$ = media de todas las observaciones, y N = número de observaciones.

booksmedicos.org

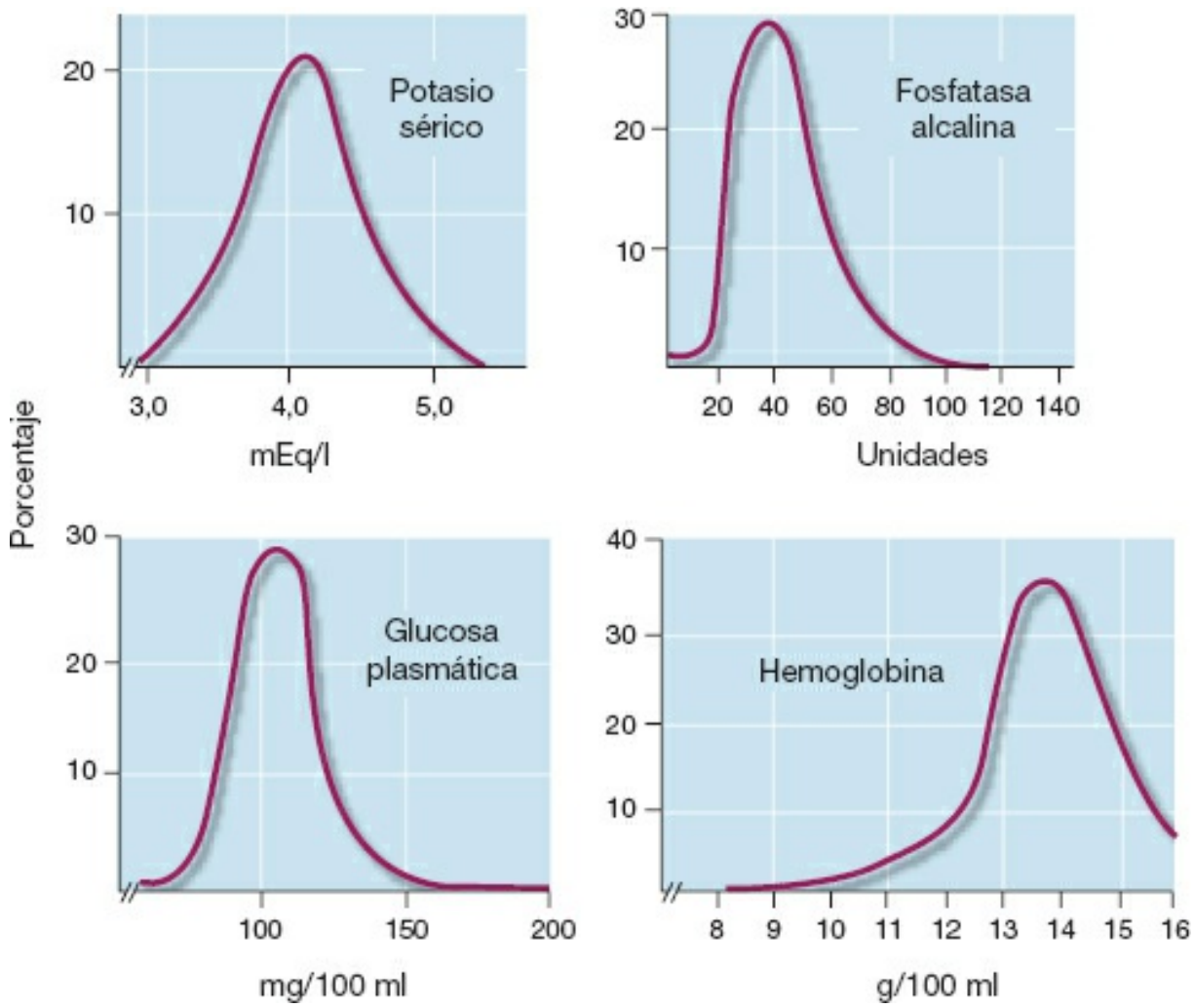


Figura 3-6 ■ Distribuciones clínicas reales. (De Martin HF, Gudzinowicz BJ, Fanger H. Normal Values in Clinical Chemistry. New York: Marcel Dekker; 1975.)

Distribución normal

Otro tipo de distribución es la denominada **distribución normal** (o «gaussiana», bautizada así en honor de Gauss, el matemático que la describió por primera vez). Esta distribución, basada en la teoría estadística, describe la distribución de frecuencia de repetidas mediciones del mismo objeto físico realizadas por el mismo instrumento. La dispersión de los valores se debe únicamente a la variación aleatoria. En la figura 3-7 se muestra una curva normal. . La curva es simétrica y con forma de campana. Esta curva tiene la propiedad matemática de que unos 2/3 de las observaciones se sitúan dentro de una desviación estándar de la media, y alrededor del 95 %, dentro de dos desviaciones estándar.

booksmedicos.org

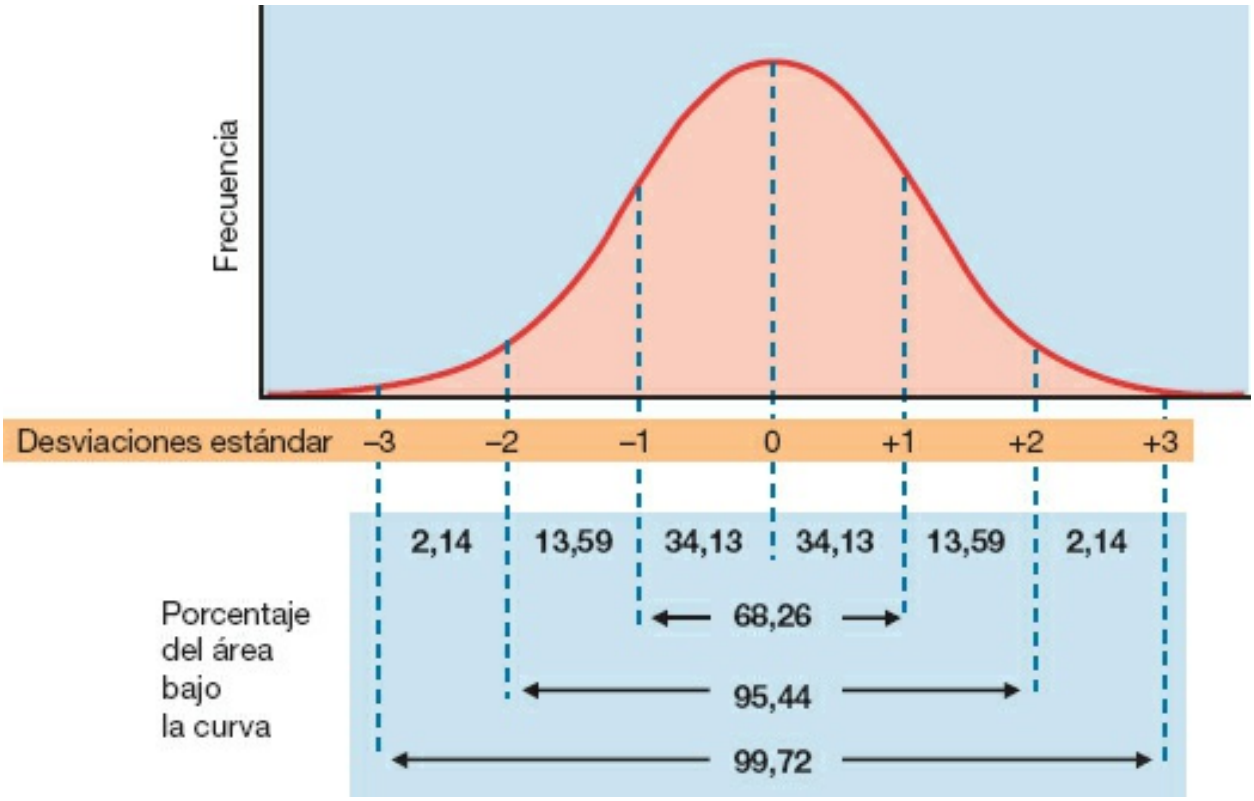


Figura 3-7 ■ Distribución normal (gaussiana).

Aunque a menudo las distribuciones clínicas se parecen a las distribuciones normales, el parecido es superficial. Como se resume en la perspectiva: «El hecho experimental es que para la mayor parte de las variables fisiológicas, la distribución es uniforme, unimodal y sesgada (asimétrica), y que la media ± 2 desviaciones estándar no se puede utilizar como punto de corte para el 95 % deseado. No disponemos de teoremas matemáticos, estadísticos o de otro tipo, que nos permitan predecir la forma que adoptarán las distribuciones de las mediciones fisiológicas» (5).

Las formas de las distribuciones clínicas difieren entre sí porque las numerosas diferencias entre las personas, además de la variación aleatoria, contribuyen a las distribuciones de las mediciones clínicas. Aun así, a menudo se supone, por comodidad (puesto que las medias y las desviaciones estándares son relativamente fáciles de calcular y de manipular de forma matemática), que las mediciones clínicas se distribuyen de forma «normal».

CRITERIOS DE ANORMALIDAD

Resultaría práctico que las distribuciones de frecuencia de las mediciones

booksmedicos.org

clínicas realizadas en pacientes sanos y en pacientes enfermos fueran tan diferentes que pudieran utilizarse para distinguir dos poblaciones diferentes. Este es realmente el caso de algunos genes anormales. Las secuencias (que codifican anomalías genéticas) para la enfermedad autosómica dominante poliposis adenomatosa familiar pueden estar presentes o ausentes. Las personas que poseen el gen anómalo presentarán cientos de pólipos en el colon, mientras que las personas sin dicho gen rara vez tienen alguno. Sin embargo, esta es la excepción que confirma la regla. Lo más habitual es que las diferentes anomalías genéticas que codifican la misma enfermedad produzcan toda una gama de expresiones. Incluso la expresión de una anomalía genética específica (como la sustitución de un solo par de bases) difiere sustancialmente de una persona a otra, probablemente debido a diferencias en el resto de la dotación genética y a la exposición a las causas externas de la enfermedad.

Por lo tanto, la mayor parte de las distribuciones de variables clínicas no se dividen tan fácilmente en «normales» y «anormales». No son inherentemente dicotómicas ni muestran brechas pronunciadas o dos picos que caractericen los resultados normales y anormales. Esto es así porque la enfermedad suele adquirirse por grados, de forma que se produce una suave transición entre los valores bajos y los valores elevados, con grados de disfunción cada vez mayores. Los resultados de las pruebas de laboratorio que reflejan insuficiencias orgánicas, como la creatinina sérica para la insuficiencia renal o la fracción de eyección para la insuficiencia cardíaca, se comportan de este modo.

Otra razón por la que pacientes sanos y enfermos no se aprecian como distribuciones separadas es que, incluso cuando las personas con una enfermedad y las que no la sufren, presentan distribuciones de frecuencia sustancialmente diferentes, dichas distribuciones casi siempre se superponen. Cuando ambas distribuciones se mezclan, como sucede en las poblaciones que existen de modo natural, los valores anormales no suelen percibirse de forma separada, debido a que ocupan una proporción muy pequeña del conjunto. La curva de las personas enfermas es «engullida» por la curva de mayor tamaño de las personas sanas.

Ejemplo

booksmedicos.org

La fenilcetonuria (PKU) es una enfermedad congénita caracterizada por un retraso mental progresivo durante la infancia. Diversos alelos mutantes que codifican la fenilalanina hidroxilasa provocan la disfunción de la enzima y, con una alimentación normal, se produce una acumulación de fenilalanina que termina causando síntomas. En principio, los genes pueden ser anormales o no (fig. 3-8 A). El diagnóstico, que se manifiesta durante el primer año de vida, lo confirman la persistencia de elevados niveles de fenilalanina (varias veces los valores habituales) y los bajos niveles de tirosina en la sangre.

Es una práctica habitual realizar pruebas de detección de PKU a los neonatos mediante análisis de fenilalanina en sangre a los pocos días de nacer, a tiempo para recibir tratamiento antes de que los daños sean irreversibles. Sin embargo, aunque los genes anómalos estén presentes o ausentes (fig. 3-8 A), la prueba clasifica de forma errónea a algunos lactantes debido a que, a esa edad, las distribuciones de las concentraciones séricas de fenilalanina en los niños con y sin PKU se superponen, y a que los lactantes con PKU constituyen sólo una pequeña proporción de todos los que se someten a la prueba de detección, es decir, menos de 1/10 000 (fig. 3-8 B). Algunos recién nacidos con PKU se sitúan dentro del intervalo normal, ya sea porque aún no han ingerido suficientes proteínas que contengan fenilalanina, o bien porque presentan una combinación de alelos que se asocia a una manifestación leve de la enfermedad. Algunos niños que no están destinados a presentar PKU tienen unos niveles de fenilalanina relativamente altos, por ejemplo, porque sus madres sufren anomalías en el metabolismo de este aminoácido. El resultado de la prueba se ha establecido como positivo en el extremo inferior de la superposición, entre los niveles normales y anormales, a fin de detectar a la mayoría de los lactantes con la enfermedad, a pesar de que sólo 1 de cada 10 lactantes con resultados anormales en la prueba de tamizaje estará realmente afectado por la PKU.

booksmedicos.org

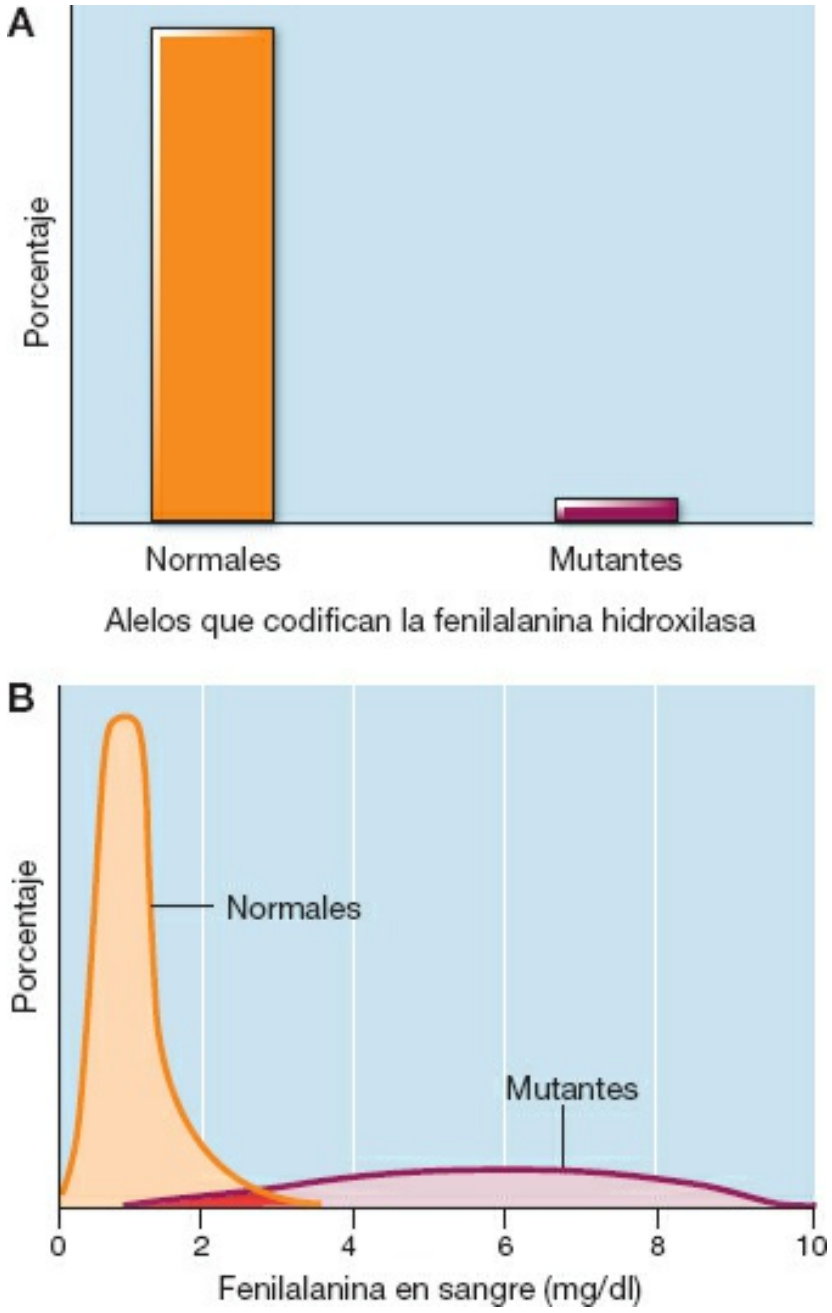


Figura 3-8 ■ Detección de fenilcetonuria (PKU) en lactantes: distribuciones dicotómicas y superpuestas de lo normal y lo anormal. A. Los alelos que codifican la fenilalanina hidroxilasa son normales o mutantes. **B.** Las distribuciones de los niveles de fenilalanina en sangre en los neonatos con y sin PKU se superponen, y la magnitud de ambas es considerablemente diferente. (La prevalencia de la PKU, que es inferior a 1/10 000, se ha exagerado, a fin de que pueda observarse su distribución en la figura.)

Si no existe una clara línea divisoria entre lo normal y lo anormal, y el médico puede elegir dónde se sitúa dicha línea, ¿qué normas básicas deben utilizarse para decidir? Existen tres criterios que han demostrado ser útiles:

booksmedicos.org

que se trate de un hecho infrecuente, que se asocie a un estado de enfermedad y que sea tratable. Para una medición concreta, los resultados de estos métodos no tienen por qué estar relacionados entre sí; lo que un criterio considera normal, puede que otro lo considere anormal.

Anormal = Infrecuente

El término normal a menudo se refiere a una situación que se produce con frecuencia o que es habitual. Lo que ocurre a menudo se considera normal y lo que se produce de forma infrecuente se considera anormal. Esta es una definición estadística, basada en la frecuencia de una característica en una población definida. Habitualmente, la población de referencia se compone de personas sanas, aunque no tiene por qué ser así. Por ejemplo, podemos decir que es normal presentar dolor después de una intervención quirúrgica, o prurito si se tiene un eccema.

Resulta tentador tratar de lograr un mayor grado de especificidad definiendo lo que es anormal mediante términos matemáticos. Una forma habitual de establecer un punto de corte entre lo normal y lo anormal es acordar, de forma algo arbitraria, que todos los valores situados más allá de dos desviaciones estándar de la media son anormales. Suponiendo que la distribución en cuestión se aproxima a una distribución normal, el 2,5 % de las observaciones se situarían en cada cola de la distribución, y se considerarían anormalmente elevadas o bajas.

Es obvio, como ya se ha señalado, que la mayor parte de las mediciones biológicas no producen distribuciones normales, por lo que es mejor describir los valores inusuales, independientemente, o cualquiera que sea la proporción elegida, como una fracción (o percentil) de la distribución real. De esta forma, se puede realizar una afirmación directa sobre el grado de infrecuencia de un valor sin hacer suposiciones sobre la forma de la distribución de la que procede.

A pesar de esto, la definición estadística de normalidad, con el punto de corte a dos desviaciones estándar de la media, es la más comúnmente utilizada. Sin embargo, puede ser ambigua o engañosa por varias razones:

Si todos los valores, situados más allá de un límite estadístico arbitrario, supongamos el percentil 95, se consideraran anormales, la frecuencia de todas las enfermedades sería la misma (si consideráramos que la distribución es normal, estaríamos hablando del 2,5 % de los valores,

booksmedicos.org

teniendo en cuenta sólo los extremos elevado y bajo de la distribución). Y todo el mundo sabe que la frecuencia de las enfermedades es variable; la diabetes y la artrosis son mucho más frecuentes que la ovalocitosis y la tricoleucemia.

No existe una relación general entre el grado de infrecuencia estadística y la enfermedad clínica. La relación es específica de la enfermedad en cuestión y del entorno. Por ejemplo, la obesidad es bastante habitual en EE.UU., pero infrecuente en muchos de los países subdesarrollados. Para algunas mediciones, las desviaciones de los valores habituales sólo se asocian a un grado importante de enfermedad en los valores extremos, más allá del percentil 95 o incluso el percentil 99. La insuficiencia de órganos como el hígado y los riñones sólo produce síntomas cuando se ha perdido la mayor parte de su función.

En ocasiones, los valores extremos pueden resultar, en realidad, beneficiosos. Por ejemplo, las personas con una presión arterial anormalmente baja presentan de promedio un riesgo menor de sufrir enfermedades cardiovasculares que las personas con presiones arteriales más elevadas y más habituales. Las que tienen una densidad ósea anormalmente elevada poseen de promedio un riesgo menor de sufrir fracturas.

Muchas mediciones están relacionadas con el riesgo de enfermedad en un amplio intervalo de valores, sin que exista un umbral que separe el riesgo normal de un riesgo mayor. La presión arterial es un ejemplo de ello.

Ejemplo

La figura 3-9 muestra que una presión arterial sistólica más elevada de lo habitual está relacionada con mayores tasas de mortalidad por cardiopatía isquémica en un amplio intervalo de valores, que oscilan entre 115 y 185 mm Hg. La figura muestra que incluso una presión sistólica de 120 mm Hg presenta más riesgo que 115 mm Hg, pero rara vez se prescriben medicamentos para la hipertensión en este nivel. Por cada aumento de 20 mm Hg en la presión arterial sistólica, se dobla el índice de fallecimientos por cardiopatía isquémica, sin indicios de que exista un umbral por debajo de 115 mm Hg, el menor valor para el que existen datos suficientes (6). Al examinar la pendiente de la curva, parece improbable que incluso una presión sistólica de 115 mm Hg sea absolutamente «segura». Entonces, ¿en qué punto se debe definir la presión arterial sistólica normal y anormal? En términos de tratamiento farmacológico, la meta actual se ha fijado arbitrariamente en 140 mm Hg para la mayoría de los pacientes.

booksmedicos.org

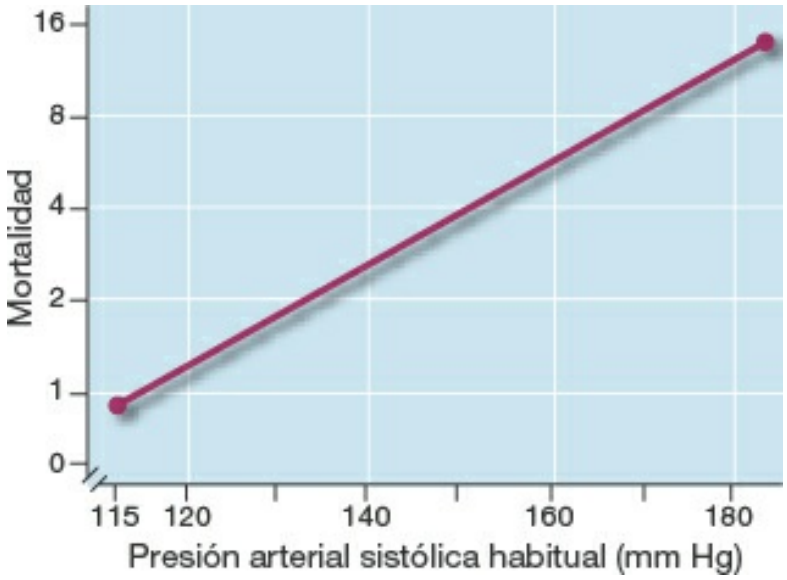


Figura 3-9 ■ La mortalidad por cardiopatía isquémica de las personas de entre 40 y 49 años de edad está relacionada con los valores de presión arterial sistólica situados a lo largo del intervalo en el que se emplazan la mayoría de las personas. No existe un umbral entre lo normal y lo anormal, ni para las personas de mediana edad ni para los ancianos. La «mortalidad» se presenta en forma de progresión geométrica respecto a la situación basal. (De Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 2002;360:1903-1913.)

Anormal = Asociado a enfermedad

Como muestra el ejemplo anterior, un método más sólido para distinguir lo normal de lo anormal consiste en considerar anómalas todas aquellas observaciones que se apartan de forma clínicamente significativa de un buen estado de salud, es decir, que están asociadas al riesgo significativo de sufrir o desarrollar una enfermedad, una discapacidad o la muerte. El uso de este enfoque para definir anormalidad a veces puede conducir a diferentes niveles en los que una enfermedad puede ser anormal.

Ejemplo

¿A partir de qué punto un peso superior a la media con respecto a la altura se convierte en un problema de salud? La figura 3-10 A muestra la relación entre el índice de masa corporal (IMC, que se obtiene dividiendo el peso de una persona, en kilogramos, por su talla, en metros, elevada al cuadrado) y el riesgo de muerte por todas las causas durante 14 años en hombres estadounidenses de 65 años o mayores que son atendidos por el seguro médico del estado (Medicare) (7). Los hombres con el menor IMC tuvieron un riesgo elevado (probablemente porque la pérdida de peso se asociaba con una enfermedad), pero el riesgo se mantuvo estable

booksmedicos.org

durante un amplio intervalo del IMC. Entre los que tenían mayor peso, sólo en aquellos con un IMC muy elevado, de 35 o más, aumentó la mortalidad. Compare este resultado con el riesgo de deterioro funcional (en la capacidad de realizar tareas domésticas sencillas y pesadas, preparar sus alimentos, hacer compras, usar el teléfono y manejar el dinero) durante 2 años (fig. 3-10 B). El porcentaje de hombres con deterioro funcional se relacionó con el peso en todo el intervalo del IMC hasta llegar a casi al doble, del 21-38%, a medida que incrementó el IMC. El punto de corte entre el IMC normal y anormal depende del resultado de salud de interés.

Anormal = Tratar el trastorno conduce a un mejor resultado clínico

Tiene sentido guiarse por la intuición de definir una enfermedad clínica o hallazgo como «anormal» si al tratarla mejora el resultado. Este enfoque tiene sentido particularmente en los trastornos asintomáticos. Si el trastorno no está causando ningún problema, y el tratamiento no hace ninguna diferencia, ¿por qué tratarlo? Sin embargo, incluso en los pacientes sintomáticos a veces es difícil distinguir entre los hallazgos clínicos que mejorarán o no con el tratamiento. La tecnología moderna, sobre todo las nuevas técnicas de imagen, ahora son tan buenas para detectar anormalidades en los pacientes que no siempre está claro si el hallazgo se relaciona con la queja del paciente. Este resultado es un dilema cada vez más habitual tanto para los pacientes como para los médicos.

booksmedicos.org

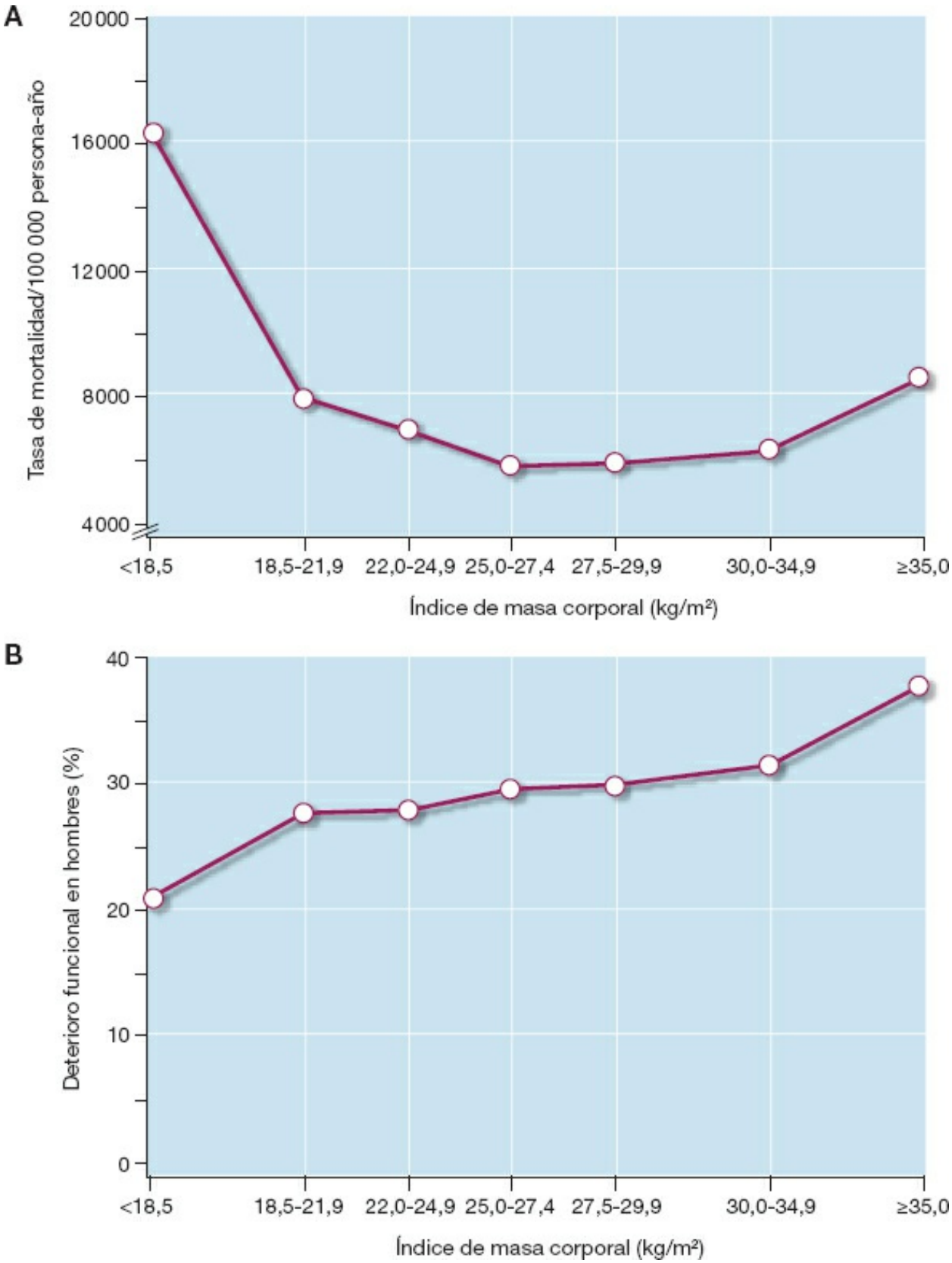


Figura 3-10 ■ Asociación de anormalidad con enfermedad y otros resultados en los pacientes.

booksmedicos.org

Relación entre el IMC y **(A)** la mortalidad total y **(B)** el deterioro funcional en hombres de 65 años o mayores tratados por el seguro médico del estado (Medicare). El IMC se calcula dividiendo el peso en kilogramos por la talla en metros elevada al cuadrado. Las tasas de mortalidad están ajustadas de acuerdo con la edad y al tabaquismo. (Redibujado con autorización de Wee CC, Huskey KW, Ngo LH, et al. Obesity, race and risk for death or functional decline among Medicare beneficiaries. A cohort study. Ann Intern Med 2011;154:645-655.)

Ejemplo

La resonancia magnética (RM) de rodilla se realiza con frecuencia a hombres de mediana edad y a pacientes de edad avanzada que tienen dolor en la rodilla; si se encuentra un desgarro en los meniscos, generalmente los síntomas se atribuyen al desgarro y es posible que se practique una meniscectomía artroscópica. No obstante, la prevalencia de los desgarros en los meniscos y su relación con los síntomas en la rodilla no está clara, sobre todo en pacientes ancianos. Se llevó a cabo un estudio a 991 adultos de mediana edad y adultos mayores seleccionados al azar, sin una cirugía previa de rodilla, a los que se les realizó una RM de esa región y se les interrogó (de manera independiente) sobre los síntomas en dicha articulación (8). Entre las personas con evidencia radiográfica de osteoartritis, el 63% de aquellos con dolor frecuente en la rodilla, molestias o rigidez tuvieron al menos un desgarro en los meniscos evidenciado por la RM, un resultado que podría haberse sospechado. De forma sorprendente, el resultado fue casi el mismo (60%) entre la personas sin estos síntomas. Entre los participantes sin evidencia de osteoartritis, los porcentajes comparables fueron 32% y 23%. Los autores concluyeron que los desgarros en los meniscos son frecuentes, independientemente de los síntomas en la rodilla y a menudo acompañan a la osteoartritis de rodilla. Irónicamente, el tratamiento quirúrgico para la reparación de meniscos aumenta el riesgo de osteoartritis, elevando la complejidad para determinar si un desgarro en el menisco es de hecho la causa del dolor en la rodilla de un paciente.

Otra de las razones para limitar la definición de «anormal» a un trastorno que se puede tratar es que no todos los trastornos que confieren un riesgo mayor pueden ser tratados con éxito: puede que la eliminación del proceso no elimine el riesgo, ya sea porque el propio proceso no es la causa de la enfermedad, sino que sólo está relacionado con la causa, o porque ya se ha producido un daño irreversible. Etiquetar a una persona de «enfermo» puede provocar preocupación y un sentimiento de vulnerabilidad que no están justificados si el tratamiento no puede mejorar el pronóstico.

Lo que se considera tratable cambia con el tiempo. En el mejor de los casos, las decisiones terapéuticas se basan en pruebas procedentes de ensayos clínicos bien diseñados (v. cap. 9). A medida que se adquieren nuevos conocimientos derivados de los resultados de los ensayos clínicos, cambia el nivel a partir del cual el tratamiento se considera beneficioso.

booksmedicos.org

Ejemplo

El ácido fólico, una vitamina que se encuentra principalmente en las verduras, fue descubierto a principios del siglo pasado. Se observó que las dietas con un bajo contenido de ácido fólico provocaban un síndrome de hipovitaminosis caracterizado por la aparición de anemia. Durante muchos años, el consumo suficiente de ácido fólico se definió en los estándares RDA *(recommended dietary allowance)* como la cantidad necesaria para evitar la anemia. Sin embargo, recientemente, han surgido argumentos a favor de unos nuevos niveles de consumo «normal» más elevados. Los niveles bajos de ácido fólico en el embarazo temprano estaban vinculados a defectos del tubo neural en los recién nacidos, a pesar de que los niveles de ácido fólico se encontraban muy por encima de los necesarios para prevenir la anemia. Los estudios subsecuentes mostraron que los suplementos de ácido fólico en mujeres de alto riesgo previenen cerca del 75% de las malformaciones. Por lo tanto, se ha establecido que la ingesta «normal» de ácido fólico durante los años fértiles debe ser mayor de acuerdo con la nueva información sobre los niveles necesarios para prevenir la enfermedad. El nivel óptimo es, al menos, dos veces (y algunos sugieren que hasta ocho veces) superior al antiguo criterio de aporte normal. Por una razón práctica, el consumo de tales niveles sólo resulta factible, para la mayoría de las personas, mediante aportes vitamínicos, y no sólo con la alimentación.

REGRESIÓN A LA MEDIA

Cuando los médicos se encuentran con un resultado anormal en una prueba, tienden a repetirla. A menudo, el resultado de la segunda prueba se acerca más a los valores normales. ¿Por qué sucede esto? ¿Debería ser tranquilizador?

Se puede esperar que los pacientes seleccionados por representar un valor extremo en una distribución tendrán de promedio menos valores extremos en las mediciones posteriores. Este fenómeno, denominado **regresión a la media,** se produce por puras razones estadísticas, y no porque los pacientes hayan mejorado necesariamente.

El origen de la regresión a la media es el siguiente (fig. 3-11): las personas son seleccionadas para su inclusión en un estudio o para un posterior diagnóstico o tratamiento, porque la medición inicial de una de sus características se sitúa más allá de un punto de corte seleccionado arbitrariamente en la cola de una distribución de valores que pertenece a un conjunto de pacientes examinados. En mediciones posteriores, algunas de estas personas seguirán por encima del punto de corte, debido a que sus valores reales suelen ser más elevados que la media. Pero otras, cuyos valores se situaron por encima del punto de corte durante el tamizaje inicial, suelen

booksmedicos.org

presentar valores inferiores. Se seleccionaron únicamente porque dio la casualidad, originada por la variación aleatoria, de que presentaron un valor elevado en el momento de la primera medición. Al repetir la medición, estas personas presentan valores inferiores a los obtenidos en el primer tamizaje. Este fenómeno tiende a producir el descenso del valor medio del subgrupo que originalmente presentó valores situados por encima del punto de corte.

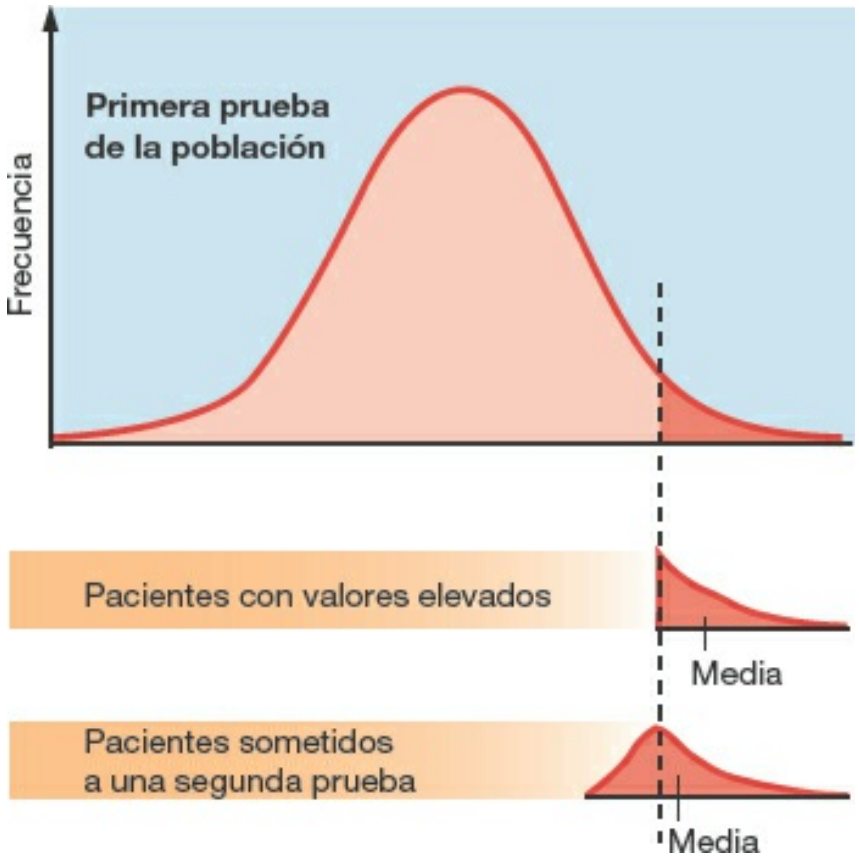


Figura 3-11 ■ Regresión a la media.

Así pues, puede esperarse que aquellos pacientes seleccionados por el resultado anormalmente elevado o bajo de una prueba analítica se acercarán de promedio al centro de la distribución, si se repite dicha prueba. Asimismo, es posible que los valores posteriores constituyan estimaciones más exactas del valor real, que podría obtenerse si las mediciones se repitieran muchas veces en un paciente concreto. De este modo, la práctica tradicional de repetir las pruebas analíticas con resultados anómalos y de considerar que el segundo es el correcto no es tan sólo cuestión de pensamientos bienintencionados, pues cuenta con el respaldo de una base teórica sólida y también de una base

booksmedicos.org

empírica. Por ejemplo, en un estudio sobre las pruebas de funcionamiento hepático (aspartato aminotransferasa, alanina aminotransferasa, fosfatasa alcalina, γ-glutamiltransferasa y bilirrubina) en una muestra representativa de la población de EE.UU., el 12-38 % de los participantes que tenían valores anormalmente altos en las pruebas iniciales tuvieron valores normales al repetir las pruebas (9). Sin embargo, cuanto más extremo es el valor inicial, más del doble del valor normal, es más probable que continúe siendo anormal cuando se repita la prueba. En los participantes con valores normales al inicio, los valores se mantuvieron dentro de la normalidad al repetir la prueba en más del 95 % de los casos.

Preguntas de revisión

Para cada uno de los escenarios clínicos que se citan a continuación (3-1 a 3-5), seleccione el término más adecuado para definir el tipo de datos a que hacen referencia.

3-1. Reflejo tendinoso profundo, grado 0 (sin respuesta), 1+ (ligeramente disminuido), 2+ (normal), 3+ (más activo que la media) y 4+ (muy activo).

A. De intervalo: continuos.
B. Dicotómicos.
C. Nominales.
D. Ordinales.
E. De intervalo: discretos.

3-2. Cáncer recurrente/no recurrente 5 años después del tratamiento inicial.

A. De intervalo: continuo.
B. Dicotómico.
C. Nominal.
D. Ordinal.
E. De intervalo: discreto.

3-3. Sodio sérico de 139 mg/dl.

booksmedicos.org

A. De intervalo: continuo.
B. Dicotómico.
C. Nominal.
D. Ordinal.
E. De intervalo: discreto.

3-4. Tres crisis convulsivas al mes.

A. De intervalo: continuo.
B. Dicotómico.
C. Nominal.
D. Ordinal.
E. De intervalo: discreto.

3-5. Causas de hemorragia digestiva alta: úlcera duodenal, gastritis, varices esofágicas o de otro tipo.

A. De intervalo: continuo.
B. Dicotómico.
C. Nominal.
D. Ordinal.
E. De intervalo: discreto.

Elija la mejor respuesta para las preguntas 3-6 a 3-10.

3-6. Cuando no es posible comprobar la medida de un fenómeno, como el prurito, mediante los sentidos físicos, ¿cuál de las siguientes afirmaciones es aplicable a su validez?

A. Es dudosa y hay que confiar en las medidas «duras», como las pruebas de laboratorio.
B. Su validez puede comprobarse demostrando que, cuando observadores diferentes repiten la medición en momentos diferentes, obtienen el mismo valor.
C. Su validez puede respaldarse si se demuestra que la medición está relacionada con mediciones de otros fenómenos, como la presencia de enfermedades que se sabe que provocan prurito.
D. Su validez puede comprobarse demostrando que la medición da

booksmedicos.org

lugar a un amplio intervalo de valores.

E. Su validez no puede establecerse.

3-7. Un médico o una enfermera miden la frecuencia cardíaca de una paciente tomándole el pulso durante 10 s cada vez que esta acude a la consulta. Las frecuencias obtenidas pueden diferir de una visita a otra debido a todas las razones citadas a continuación excepto:

A. La paciente presenta diferentes frecuencias de pulso en momentos diferentes.

B. La medición puede representar de forma errónea el pulso verdadero por efecto del azar y debido al breve período de observación.

C. El médico y la enfermera utilizan técnicas diferentes (p. ej., diferentes grados de presión sobre el pulso).

D. La frecuencia del pulso varía entre los pacientes.

E. Se empezó un tratamiento eficaz entre una visita y otra.

3-8. Habitualmente, el término «anormal» se corresponde con todas las definiciones siguientes, *excepto:*

A. Nivel a partir del cual el tratamiento ha demostrado ser eficaz.

B. Nivel a partir del cual aumenta la tasa de mortalidad.

C. Valores estadísticamente infrecuentes.

D. Valores que no corresponden a una distribución normal.

E. Nivel a partir del cual existe un mayor riesgo de presentar síntomas.

3-9. Las siguientes afirmaciones son verdaderas excepto:

A. La distribución normal describe la distribución de los fenómenos que se producen de forma natural.

B. La distribución normal incluye un 2,5 % de personas en cada cola de la distribución (más allá de dos desviaciones estándar de la media).

C. La distribución normal es unimodal y simétrica.

D. La distribución normal es la mejor base para definir los resultados de laboratorio anómalos medidos en una escala de intervalo.

booksmedicos.org

3-10. Usted está tratando a un paciente nuevo, una mujer de 71 años que no toma medicamentos y no tiene antecedentes personales ni familiares de enfermedad cardíaca. Nunca ha fumado y no es diabética. Su presión arterial es de 115/75 mm Hg y tiene unos 7 kg de sobrepeso. Hace 2 días se realizó una prueba de colesterol que estaba elevada, en 250 mg/dl, mientras que el colesterol HDL era de 59 mg/dl. La calculadora de riesgo de Framingham estima que el riesgo de la paciente de desarrollar enfermedad cardiovascular general en los siguientes 10 años es del 9 %. Usted sabe que tratar el colesterol a este nivel reduce el riesgo cardiovascular. La paciente desea saber si debe comenzar a tomar una estatina. ¿Cuál de las siguientes afirmaciones es la *menos correcta*?

A. Es probable que la paciente tenga un colesterol sérico más bajo la próxima vez que sea medido.

B. La estimación sobre la probabilidad de enfermedad cardiovascular en los siguientes 10 años del 9 % puede estar influida por el azar.

C. Debe prescribir a la paciente estatinas para disminuir el riesgo de enfermedad cardíaca coronaria.

La figura 3-12 muestra la distribución de peso al nacer de más de un millón de bebés nacidos a término de madres sin diabetes. Las preguntas 3-11 a 3-13 se relacionan con la figura. Elija la mejor respuesta para cada pregunta.

booksmedicos.org

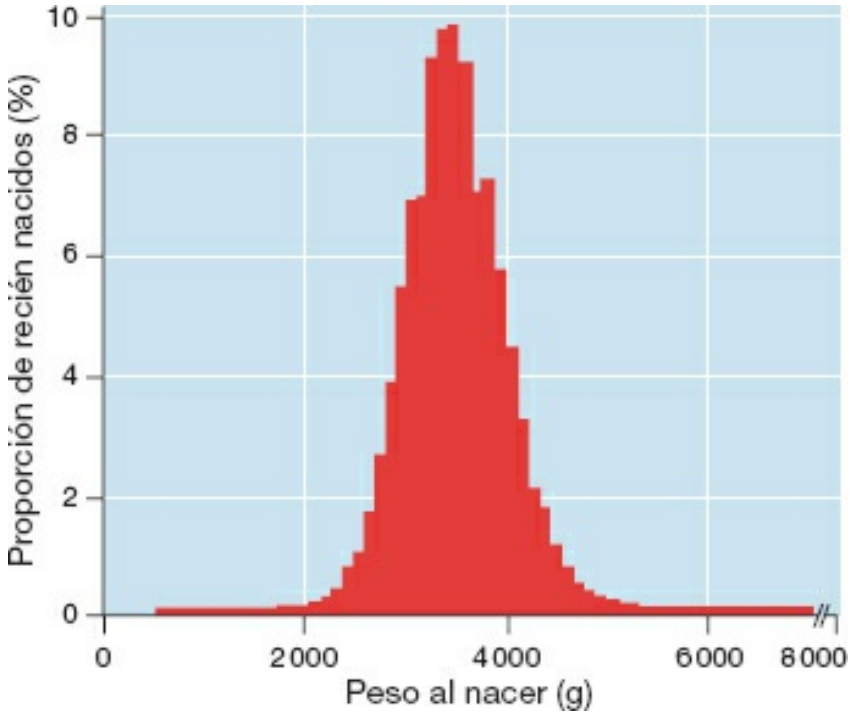


Figura 3-12 ■ La distribución de los pesos al nacer de los bebés a término de madres sin diabetes. (Redibujado con autorización de Ludwig DS, Currie J. The association between pregnancy weight gain and birthweight: a within-family comparison. Lancet 2010;376:984-990.)

3-11. ¿Qué afirmación sobre la tendencia central es *incorrecta?*

A. El peso promedio al nacer está por debajo de 4 000 g.
B. Existe más de una moda para el peso al nacer.
C. La media y la mediana del peso al nacer son similares.

3-12. ¿Qué afirmación sobre la dispersión es la más correcta?

A. El intervalo es la mejor manera de describir el peso al nacer de los recién nacidos.
B. No deben calcularse las desviaciones estándar porque la distribución está sesgada.
C. El 95 % de los pesos al nacer disminuirán o quedarán comprendidos dentro de 2 desviaciones estándar respecto a la media.

3-13. Una desviación estándar de los pesos al nacer de los recién nacidos comprende aproximadamente:

A. Pesos de 2 000 a 4 000 g.
B. Pesos de 3 000 a 4 000 g.

booksmedicos.org

C. Pesos de 2 000 a 6 000 g.

La figura 3-13 compara la frecuencia cardíaca fetal medida con la monitorización electrónica (la barra blanca intermedia) con las mediciones del personal hospitalario en tres circunstancias diferentes: cuando la frecuencia cardíaca fetal en latidos por minuto medida mediante monitorización electrónica era normal (130-150), baja (< 130) y alta (> 150). Las preguntas 3-14 a 3-16 se relacionan con la figura. Elija la mejor respuesta para cada pregunta.

booksmedicos.org

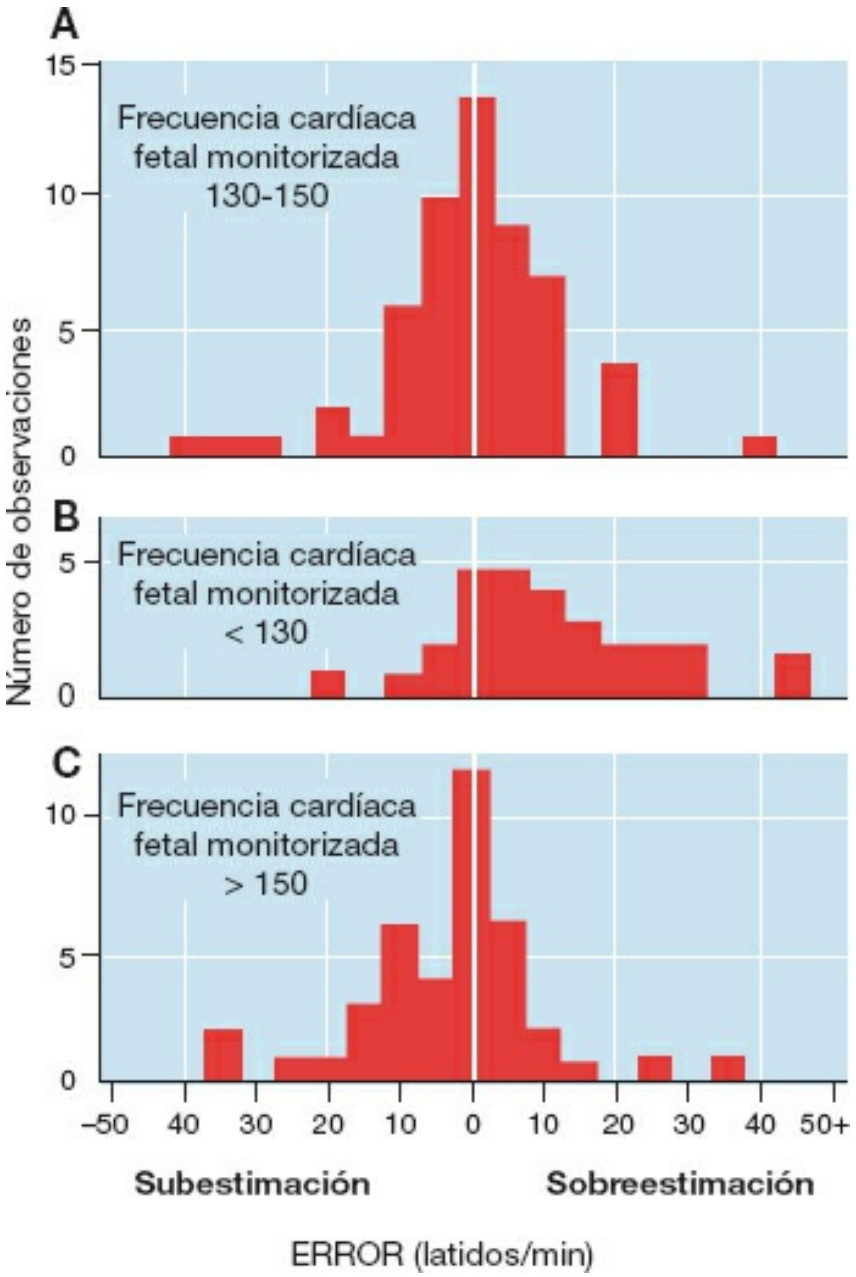


Figura 3-13 ■ **Variabilidad del observador.** Comparación de la auscultación cardíaca fetal con la monitorización electrónica de la frecuencia cardíaca fetal. (Redibujado con autorización de Day E, Maddem L, Wood C. Auscultation of foetal heart rate: an assessment of its error and significance. Br Med J 1968;4:422-424.)

3-14. La distribución de las mediciones del personal del hospital en torno a las mediciones del monitor electrónico en la figura 3-13 A podrían deberse a:

A. Casualidad.
B. Variabilidad interobservador.

booksmedicos.org

C. Preferencia sesgada por los resultados normales.
D. A y B.
E. A y C.
F. B y C.
G. A, B y C.

3-15. La distribución de las mediciones del personal del hospital en torno a las mediciones del monitor electrónico en la figura 3-13 B podrían deberse a:

A. Casualidad.
B. Variabilidad interobservador.
C. Preferencia sesgada por los resultados normales.
D. A y B.
E. A y C.
F. B y C.
G. A, B y C.

3-16. La distribución de las mediciones del personal del hospital en torno a las mediciones del monitor electrónico en la figura 3-13 C podrían deberse a:

A. Casualidad.
B. Variabilidad interobservador.
C. Preferencia sesgada por los resultados normales.
D. A y B.
E. A y C.
F. B y C.
G. A, B y C.

Encontrará las respuestas en el apéndice A.

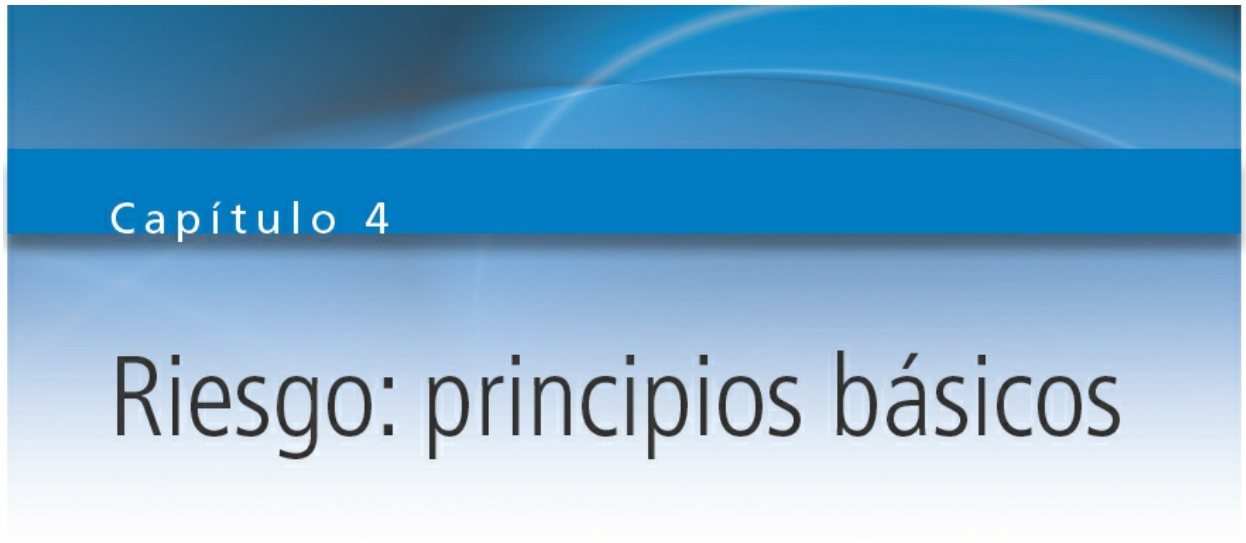
BIBLIOGRAFÍA

1. Sharma P, Parekh A, Uhl K. An innovative approach to determine fetal risk: the FDA Office of Women's Health pregnancy exposure registry web listing. Womens Health Issues 2008;18: 226–228.
2. Feinstein AR. The need for humanized science in evaluating medication. Lancet 1972;2:421–423.
3. Rubenfeld GD, Caldwell E, Granton J, et al. Interobserver variability in applying a radiographic

booksmedicos.org

definition for ARDS. Chest 1999;116:1347–1353.
4. Morganroth J, Michelson EL, Horowitz LN, et al. Limitations of routine long-term electrocardiographic monitoring to assess ventricular ectopic frequency. Circulation 1978;58:408–414.
5. Elveback LR, Guillier CL, Keating FR. Health, normality, and the ghost of Gauss. JAMA 1970;211:69–75.
6. Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 2002;360:1903–1913.
7. Wee CC, Huskey KW, Ngo LH, et al. Obesity, race and risk for death or functional decline among Medicare beneficiaries. A cohort study. Ann Intern Med 2011;154:645–655.
8. Englund M, Guermazi A, Gale D. Incidental meniscal findings on knee MRI in middle-aged and elderly persons. N Engl J Med 2008;359:1108–1115.
9. Lazo M, Selvin E, Clark JM. Brief communication: clinical implications of short-term variability in liver function test results. Ann Intern Med 2008;148:348–352.

booksmedicos.org



La lección... es que un gran número de personas con un riesgo bajo pueden originar más casos de una enfermedad que un pequeño número de personas con un riesgo alto.

—Geoffrey Rose
1985

PALABRAS CLAVE

Riesgo
Factores de riesgo
Exposición
Período de latencia
Causas inmediatas
Causas lejanas
Marcador
Modelo de predicción del riesgo
Herramienta de predicción del riesgo
Estratificación del riesgo
Calibración
Discriminación
Concordancia estadística
Estadística c
Sensibilidad
Especificidad
Curva característica operativa del receptor (ROC)

En general, la idea de **riesgo** se refiere a la probabilidad de que se produzca

booksmedicos.org

un acontecimiento desfavorable. En medicina, los médicos tratan con probabilidades en prácticamente cualquier consulta con el paciente. Trabajan con los principios básicos de riesgo si están diagnosticando una afección, describiendo un pronóstico, decidiendo sobre un tratamiento o discutiendo la prevención con el paciente. Las consultas con este ya no se tratan sólo sobre sus molestias, sino que cada vez más implican la revisión de factores de riesgo que podrían producir problemas de salud en el futuro. En la investigación médica se realizan cada vez más esfuerzos para dilucidar los factores de riesgo de las enfermedades. Una justificación importante del Proyecto del Genoma Humano es predecir las enfermedades que se manifestarán durante la vida de una persona. Las revistas médicas publican los estudios epidemiológicos que investigan los posibles riesgos para la salud. Todos estos esfuerzos han mejorado nuestra comprensión de los riesgos de la salud y cómo estudiarlos, y han ayudado a mejorar la salud de los pacientes y las poblaciones algunas veces de forma dramática. Por ejemplo, la investigación que condujo al conocimiento que el tabaquismo, la hipertensión y la hiperlipidemia aumentan el riesgo de enfermedad cardiovascular (ECV) ha ayudado a disminuir a la mitad la mortalidad cardiovascular en EE.UU. en las últimas décadas.

Las personas sienten un gran interés por su riesgo de enfermar, que se refleja en los programas de televisión y en los titulares de los periódicos, además de muchos sitios de internet y en libros populares que tratan sobre la reducción del riesgo. Algunos ejemplos de los temas sobre los que el público ha desarrollado gran interés son el riesgo de cáncer de mama y prostático, de enfermedad cardíaca y cerebrovascular, de enfermedad de Alzheimer, autismo y osteoporosis. Los pacientes más exigentes desean saber sus riesgos individuales y cómo reducirlos.

En este capítulo nos concentraremos en los principios subyacentes sobre el riesgo, relevantes con independencia de la parte del espectro de atención que afrontan el médico y el paciente. En la medida de lo posible hablaremos de conceptos en lugar de terminología y métodos, abarcando esos detalles importantes en los últimos capítulos. Los capítulos 5 y 6 tratan de los riesgos de los efectos adversos para la salud en un futuro lejano, a menudo años o incluso décadas después; se describen los métodos científicos utilizados para indicar la probabilidad de que las personas expuestas a ciertos «factores de riesgo» desarrollen después una enfermedad particular o un resultado adverso para la salud con mayor frecuencia que las personas similares que no han

booksmedicos.org

estado expuestas. En el contexto de la atención urgente, como en las unidades de cuidados intensivos, las salas de urgencias y los pabellones de hospitalización, la preocupación se basa en los riesgos que los pacientes con una enfermedad conocida podrían o no experimentar lo que se denomina pronóstico (v. cap. 7); el límite temporal del pronóstico abarca desde minutos u horas hasta meses y años, dependiendo del estudio. Los capítulos 8, 9 y 10 vuelven a examinar el riesgo en relación con el diagnóstico, tratamiento y prevención. En cada caso, el abordaje para evaluar el riesgo es algo diferente. Sin embargo, los principios fundamentales para determinar los riesgos a la salud son similares.

FACTORES DE RIESGO

Las características asociadas a un mayor riesgo de enfermar se denominan **factores de riesgo**, y de ellos, algunos son hereditarios. Por ejemplo, poseer el haplotipo HLA-B27 aumenta enormemente el riesgo de sufrir espondiloartropatías. El trabajo llevado a cabo en el Proyecto del Genoma Humano ha identificado otras muchas enfermedades para las que genes específicos constituyen factores de riesgo, como el cáncer de colon y el de mama, la osteoporosis y la esclerosis lateral amiotrófica. Otros factores de riesgo, como agentes infecciosos, drogas y toxinas se encuentran en el ambiente físico, y otros forman parte del entorno social. Por ejemplo, se ha demostrado que el duelo tras la pérdida de un cónyuge, los cambios en las rutinas diarias y las multitudes aumentan las tasas de las enfermedades (no sólo emocionales, sino también físicas). Algunos de los factores de riesgo más potentes son conductuales: fumar, ingerir bebidas alcohólicas en exceso, conducir sin cinturón de seguridad, mantener relaciones sexuales sin protección, comer demasiado y practicar poco ejercicio físico son ejemplos de ello.

La **exposición** a un factor de riesgo significa que una persona, antes de contraer la enfermedad, ha estado en contacto con el factor en cuestión o lo ha presentado. La exposición puede producirse en un punto determinado del tiempo, como cuando una comunidad queda expuesta a la radiación tras un accidente nuclear. Sin embargo, con mayor frecuencia, el contacto con los factores de riesgo de una enfermedad crónica tiene lugar a lo largo de un período de tiempo. El consumo de cigarrillos, la hipertensión arterial, la promiscuidad sexual y la exposición solar son ejemplos de factores de riesgo

booksmedicos.org

con los que el riesgo de sufrir la enfermedad se adquiere, con mayor frecuencia, tras una exposición prolongada.

Existen varias formas diferentes de caracterizar la cantidad de exposición o contacto con un supuesto factor de riesgo: exposición en alguna ocasión, dosis actual, mayor dosis recibida, dosis acumulada total, años de exposición, años transcurridos desde la primera exposición, etc. Aunque las diversas medidas de las dosis tienden a relacionarse entre sí, algunas pueden mostrar una relación entre la exposición y la enfermedad que en otras puede no apreciarse. Por ejemplo, las dosis acumuladas de exposición solar constituyen un factor de riesgo de cáncer de piel distinto al melanoma, mientras que los episodios de quemaduras solares intensas son mejores predictores de este último. Si no se elige la medida correcta, puede no apreciarse la asociación entre un factor de riesgo y una enfermedad. La elección de una medida adecuada de exposición a un factor de riesgo suele basarse en todos los conocimientos que se poseen acerca de los efectos clínicos y biológicos de la exposición, la fisiopatología de la enfermedad y los estudios epidemiológicos previos.

RECONOCER EL RIESGO

Cualquiera puede reconocer con facilidad los grandes riesgos asociados a efectos que se producen rápidamente tras la exposición. No es difícil apreciar la relación entre la exposición y algunas afecciones como la dermatitis actínica aguda y la sobredosis de ácido acetilsalicílico, o el mal pronóstico de la hipotensión al inicio de un infarto de miocardio, pues el efecto perjudicial ocurre rápidamente después de la exposición y es fácil identificarlo.

También es fácil identificar el aumento repentino de las enfermedades infrecuentes o la dramática presentación clínica de una nueva enfermedad. El SIDA fue un síndrome tan inusual que la aparición de unos pocos casos despertó la sospecha de que algún agente nuevo (como lo fue el retrovirus) podría ser responsable, lo que se confirmó relativamente rápido después de los primeros casos de la enfermedad. En 2003 se informaron de los primeros casos del síndrome respiratorio agudo del adulto (SARS) y en unas cuantas semanas se confirmó que un coronavirus previamente no identificado causaba esta infección altamente letal. De forma similar, hace algunas décadas, los médicos se percataron con rapidez de la aparición de varios casos de carcinoma vaginal, una afección muy rara. Se llevó a cabo una investigación

booksmedicos.org

detallada para buscar una explicación, que resultó ser la exposición materna a dietilestilbestrol (una hormona utilizada para estabilizar el embarazo en mujeres con antecedente de aborto).

Sin embargo, la mayor parte de la morbilidad o mortalidad está causada por enfermedades crónicas para las que la relación entre la exposición y la enfermedad no es ni mucho menos tan evidente. Suele ser imposible que un solo médico, por muy perspicaz que sea, reconozca los factores de riesgo de una enfermedad crónica, basándose en su propia experiencia con los pacientes. Esto es así por varios motivos, que se explican a continuación.

Latencia prolongada

Muchas enfermedades crónicas presentan largos **períodos de latencia** entre la exposición a un factor de riesgo y las primeras manifestaciones de la enfermedad. La exposición a la radiación durante la infancia, por ejemplo, aumenta el riesgo de cáncer tiroideo en los adultos décadas más tarde. De igual modo, la hipertensión arterial precede en decenios a la cardiopatía y el consumo de calcio en las mujeres jóvenes y de mediana edad influye en las tasas de osteoporosis y fracturas en la vejez. Cuando los pacientes sufren las consecuencias de la exposición a un factor de riesgo al cabo de los años, puede que prácticamente hayan olvidado la exposición inicial, y que el vínculo entre la exposición y la enfermedad haya quedado oculto.

Causas inmediatas frente a distantes

La búsqueda de factores de riesgo suele ser una búsqueda de las causas de una enfermedad. En medicina clínica, los médicos se sienten más cómodos con las **causas inmediatas** de enfermedad: infecciosas, fisiológicas o cambios anatómicos que dan lugar a los síntomas o signos, como el coronavirus que provoca el SARS o la hipocalcemia que da lugar a las crisis epilépticas. Sin embargo, las **causas lejanas** pueden ser importantes en la vía causal. Por ejemplo, la falta de educación materna es un factor de riesgo de tener hijos con bajo peso al nacer, aunque existen otros factores relacionados con la educación, como una nutrición deficiente, un menor cuidado prenatal o el consumo de cigarrillos, que constituyen causas más directas del bajo peso del recién nacido en el momento de nacer. Sin embargo, estudios realizados en India han mostrado que si se mejora la formación materna, se reduce la mortalidad infantil.

booksmedicos.org

Exposición habitual a factores de riesgo

Muchos factores de riesgo, como el consumo de cigarrillos o seguir una alimentación rica en azúcares, sal y grasas, se han hecho tan habituales en las sociedades industrializadas que durante muchos años apenas se han considerado peligrosos. Sólo mediante la comparación de los patrones de enfermedad entre personas con y sin estos factores de riesgo, a través de estudios transversales a nivel nacional o de la investigación de subgrupos especiales (como los mormones, que no fuman, o los vegetarianos, que siguen regímenes alimenticios bajos en colesterol), han podido identificarse riesgos que, de hecho, eran considerables. Actualmente se sabe que alrededor de la mitad de las personas que fuman durante toda su vida morirán a causa de su hábito; si se mantienen los patrones de tabaquismo actuales, se prevé que en el s. XXI más de 1 000 millones de fallecimientos a escala mundial se atribuirán al tabaco (1).

La relación entre la posición de los bebés durante el sueño y el síndrome de la muerte súbita del lactante (SMSL) es otro ejemplo de una exposición habitual a un factor de riesgo y el espectacular efecto asociado a su frecuencia, una asociación que no se detectó hasta hace poco tiempo.

Ejemplo

El SMSL, es decir, la inexplicable muerte súbita de un bebé de menos de 1 año de edad, es la principal causa de muerte infantil. Los estudios sugieren que existen muchos factores influyentes. A finales de la década de 1980 y durante la década de 1990, diversas investigaciones detectaron que era entre 3 y 9 veces más probable que los bebés que se ponían a dormir boca abajo en sus cunas murieran por SMSL que los colocados boca arriba. En 1992, la American Academy of Pediatrics (AAP) hizo pública la recomendación de colocar a los bebés en decúbito supino a la hora de dormir, pero mencionó que una alternativa aceptable era colocarlos sobre su costado. Entre 1992 y 1996, el porcentaje de bebés colocados en decúbito prono para dormir se redujo del 70 % hasta el 24 %, respectivamente, con un descenso concomitante del 40 % en el número de casos de SMSL (2). La evidencia de la investigación actual mostró que colocar a los bebés sobre su costado también aumentaba el riesgo de muerte súbita del lactante (3), por lo que la AAP actualizó sus recomendaciones en 2005 para aclarar que ya no se recomienda colocarlos sobre su costado para dormir.

Baja incidencia de enfermedad

La incidencia de la mayor parte de las enfermedades, incluso las que se

booksmedicos.org

consideran «comunes», es en realidad poco frecuente. Por lo tanto, aunque el cáncer de pulmón sea la causa más habitual de fallecimiento por cáncer entre los estadounidenses, y las personas que fuman tienen hasta 20 veces más probabilidades de desarrollarlo que aquellos que no fuman, su incidencia anual, incluso en personas que han fumado mucho durante 30 años, es de 2-3/1 000. En la práctica médica, por término medio, pueden pasar años entre los nuevos casos de cáncer de pulmón. Para el médico, es difícil extraer conclusiones sobre los riesgos de episodios tan infrecuentes.

Riesgo pequeño

Los efectos de muchos factores de riesgo son pequeños en el caso de las enfermedades crónicas. Para detectar un efecto pequeño es necesario estudiar un gran número de personas, a fin de observar alguna diferencia en las tasas de la enfermedad entre las personas que han estado expuestas y las que no. Por ejemplo, se sabe que el consumo de alcohol aumenta el riesgo de cáncer de mama, pero estaba menos claro si el consumo bajo, como tomar a diario sólo una copa de vino o su equivalente, confería algún riesgo. Era necesario un estudio de 2 400 000 años-mujeres para saber si aumentaba un 15 % el riesgo de las mujeres con el consumo promedio de una copa diaria (4). Debido al gran número de años-mujeres en el estudio, es poco probable que el azar explique el resultado, pero aun así una pequeña parte del efecto podría ser debido al sesgo. Por el contrario, no hay duda de que la infección por el virus de la hepatitis B es un factor de riesgo de sufrir hepatoma, debido a que las personas con dicha enfermedad tienen hasta 60 veces (no sólo 1,15 veces) más probabilidad de sufrir cáncer hepático que las que no la padecen.

booksmedicos.org

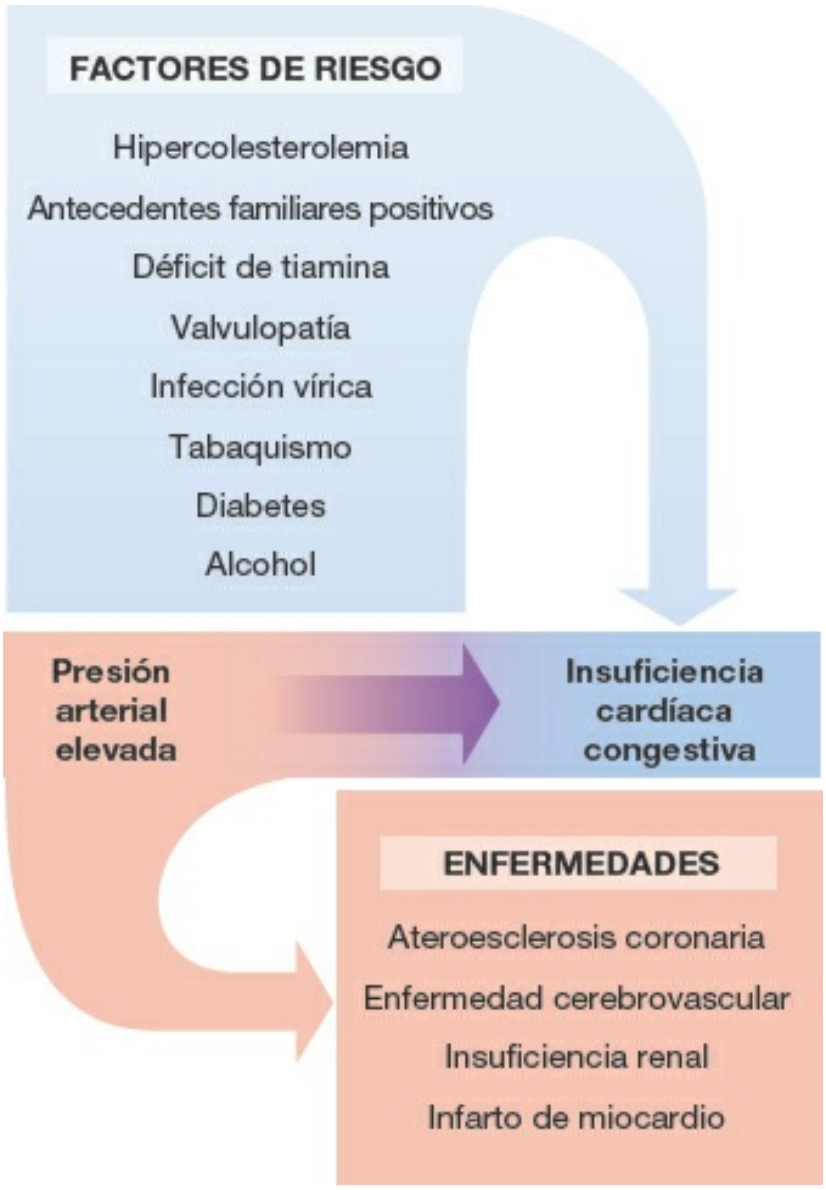


Figura 4-1 ■ Relación entre factores de riesgo y enfermedad: hipertensión arterial e insuficiencia cardíaca congestiva. La hipertensión arterial provoca muchas enfermedades, como la insuficiencia cardíaca congestiva, y la insuficiencia cardíaca congestiva está provocada por muchos motivos, entre ellos la hipertensión arterial.

Múltiples causas y múltiples efectos

Habitualmente no existe una relación estrecha y unívoca entre un factor de riesgo y una enfermedad concreta. Un determinado factor de riesgo puede contribuir a varias enfermedades y una enfermedad puede tener múltiples causas. Un ejemplo de ello es la relación entre la hipertensión arterial y la insuficiencia cardíaca congestiva (fig. 4-1). Algunas personas con hipertensión arterial presentarán insuficiencia cardíaca congestiva, y muchas

booksmedicos.org

otras no. Asimismo, muchas que no sufren hipertensión arterial presentan insuficiencia cardíaca congestiva debido a que existen otras causas. La relación también resulta difícil de detectar porque la hipertensión arterial provoca otras enfermedades, además de la insuficiencia cardíaca congestiva. Así pues, aunque la hipertensión es la tercera causa principal de insuficiencia cardíaca congestiva, los médicos no fueron conscientes de esta relación hasta la década de 1970, cuando se dispuso de suficientes indicios, tras el minucioso estudio de un gran número de personas durante muchos años.

Los factores de riesgo pueden ser o no causales

Sólo porque los factores de riesgo predigan una enfermedad no significa necesariamente que la causen. Es posible que un factor de riesgo pueda predecir un resultado de enfermedad de forma indirecta mediante la asociación con alguna otra variable que hoy en día es un determinante de enfermedad. Es decir, el factor de riesgo se confunde con un factor verdaderamente causal.

Un factor de riesgo que no constituye una causa de enfermedad se denomina **marcador** de enfermedad debido a que «marca» la mayor probabilidad de sufrirla. El hecho de que un factor de riesgo no constituya una causa no reduce su valor como factor de riesgo y como una forma de predecir la probabilidad de la enfermedad, sino que implica que la eliminación del riesgo quizá no eliminaría el exceso de riesgo asociado a él.

Ejemplo

La homocistinuria, una rara enfermedad pediátrica causada por genes autosómicos recesivos y niveles muy elevados del aminoácido homocisteína, se asocia con ateroesclerosis prematura grave. Se ha investigado la posibilidad que niveles ligeramente elevados de homocisteína también puedan ser un factor de riesgo para enfermedad cardíaca coronaria en los adultos. Múltiples estudios han encontrado que existe una relación; por cada incremento de 5 µmol/l en los niveles de homocisteína el riesgo de ECV aumenta un 20 % (5). Debido a esta relación, se llevaron a cabo estudios clínicos para determinar si la reducción de los niveles de homocisteína con ácido fólico, además de vitaminas B6 y B12, podría reducir el riesgo de episodios cardiovasculares graves entre las personas con ECV conocida o diabetes. Sorprendentemente, varios estudios grandes encontraron que, incluso disminuyendo los niveles de homocisteína, los suplementos vitamínicos no protegían contra los episodios cardiovasculares futuros en estos pacientes (6). Por lo tanto, aunque los niveles elevados de homocisteína eran un marcador para ECV futura en los adultos, no parecían ser el factor causal, al menos entre los pacientes con ECV conocida o diabetes.

booksmedicos.org

Existen varios métodos para decidir si un factor de riesgo constituye una causa o sencillamente es un marcador de enfermedad. Dichos métodos se explicarán en el capítulo 5.

Por todas estas razones, es muy raro que un médico se encuentre en posición de detectar, y mucho menos confirmar, las asociaciones entre exposición y enfermedades crónicas. Puede que perciba una asociación cuando una enfermedad llamativa se produce inmediatamente después de una exposición poco común, pero la mayoría de las enfermedades y exposiciones no se ajustan a este patrón. Para obtener una información exacta sobre el riesgo, los médicos deben recurrir a la bibliografía médica, sobre todo a estudios con diseños minuciosos en los que se ha incluido a un gran número de pacientes.

PREDICCIÓN DEL RIESGO

Un único factor de riesgo importante, como en el caso de la hepatitis B y el cáncer hepatocelular, puede ser muy útil a nivel clínico, pero la mayoría de los factores de riesgo no son fuertes. Si beber una copa de vino a diario aumenta el riesgo de cáncer de mama en un 15 %, y el riesgo promedio a 10 años de desarrollar cáncer de mama para las mujeres de 40 años es de 1 en 69 (o 1,45 %), beber una copa de vino en la cena aumentaría el riesgo a 10 años aproximadamente a 1 en 60 (o 1,67 %). Algunas mujeres pueden pensar que esta diferencia en el riesgo de cáncer de mama no es significativa.

Combinación de múltiples factores de riesgo para predecir el riesgo

Debido a que la mayoría de las enfermedades crónicas están causadas porque varios factores de riesgo relativamente débiles actúan en conjunto, la combinación estadística de sus efectos puede dar lugar a una predicción del riesgo más potente que considerando sólo un factor de riesgo a la vez. La combinación estadística de los factores de riesgo produce un **modelo de predicción del riesgo** o una **herramienta de predicción del riesgo** (a veces también denominada herramienta de predicción clínica o herramienta de evaluación del riesgo). Las herramientas de predicción del riesgo son cada vez más habituales en medicina clínica; los modelos conocidos utilizados para la predicción a largo plazo incluyen la Escala de riesgo de Framingham para predecir episodios cardiovasculares y la herramienta para la evaluación del riesgo de cáncer de mama del National Cancer Institute (NCI) para

booksmedicos.org

predecir la aparición de cáncer de mama. Las herramientas para la predicción del riesgo hospitalario a corto plazo comprenden la Puntuación de reingreso de los pacientes en riesgo (PARR, *Patient At Risk of Re-admission Scores*) y la Puntuación de alerta temprana (EWS, *Early Warning Scores*) en la atención crítica. Las herramientas de predicción también han combinado resultados de las pruebas diagnósticas, por ejemplo, para diagnosticar un infarto de miocardio cuando el paciente tiene dolor en el pecho o para diagnosticar la aparición de embolia pulmonar. Los métodos estadísticos para combinar múltiples factores de riesgo se discuten en el capítulo 11.

Los modelos de predicción del riesgo facilitan dos actividades clínicas importantes. En primer lugar, un buen modelo de predicción del riesgo ayuda a la **estratificación del riesgo**, es decir, dividir grupos de personas en subgrupos con diferentes grados de riesgo (p. ej., bajo, medio y alto). El enfoque de estratificación del riesgo también puede ayudar a determinar si la adición de un factor de riesgo recientemente propuesto para una enfermedad mejora la capacidad de predecirla, es decir, mejora la estratificación del riesgo.

Ejemplo

La ECV es la causa de muerte más habitual en el mundo. En EE.UU., a pesar de la disminución en la incidencia, la ECV representa un tercio de todas las muertes. En los últimos 50 años se han identificado varios factores de riesgo, como la hipertensión arterial, la hipercolesterolemia, el tabaquismo, la diabetes mellitus, una obesidad grave, la inactividad física y los antecedentes familiares. Cuando se utilizan en las herramientas de predicción del riesgo, estos factores ayudan a predecir el riesgo de ECV. En la práctica clínica, la predicción del riesgo es más útil cuando se estratifica a las personas en alto o bajo riesgo, para que los pacientes y sus médicos se sientan cómodos con la observación o la intervención. No está claro qué hacer con las personas con un riesgo moderado de ECV calculado con la herramienta de predicción del riesgo. Se han hecho esfuerzos para mejorar estas herramientas para ECV, añadiendo factores de riesgo que reclasifican a las personas que inicialmente tenían un riesgo moderado en categorías de alto o bajo riesgo. Uno de estos factores de riesgo es un marcador plasmático de inflamación: la proteína C reactiva (PCR). Un resumen de los estudios sobre la PCR encontró que, al añadir a los modelos de predicción del riesgo, se puede mejorar la estratificación para ECV (7). La figura 4-2 muestra que añadiendo los resultados del análisis de PCR se mejora la estratificación del riesgo en una herramienta de predicción de riesgo para ECV en mujeres, al reclasificarlas en categorías de alto o bajo riesgo, que coinciden con mayor precisión con las tasas de ECV en un período de 10 años (8).

Incluso si un factor de riesgo mejora un modelo de predicción del riesgo,

booksmedicos.org

es necesario un ensayo clínico que demuestre que la reducción o eliminación del factor de riesgo protege a los pacientes. En el caso de la PCR, no se ha comunicado tal ensayo clínico, por lo que es posible que sea un marcador en vez de un factor causal de ECV.

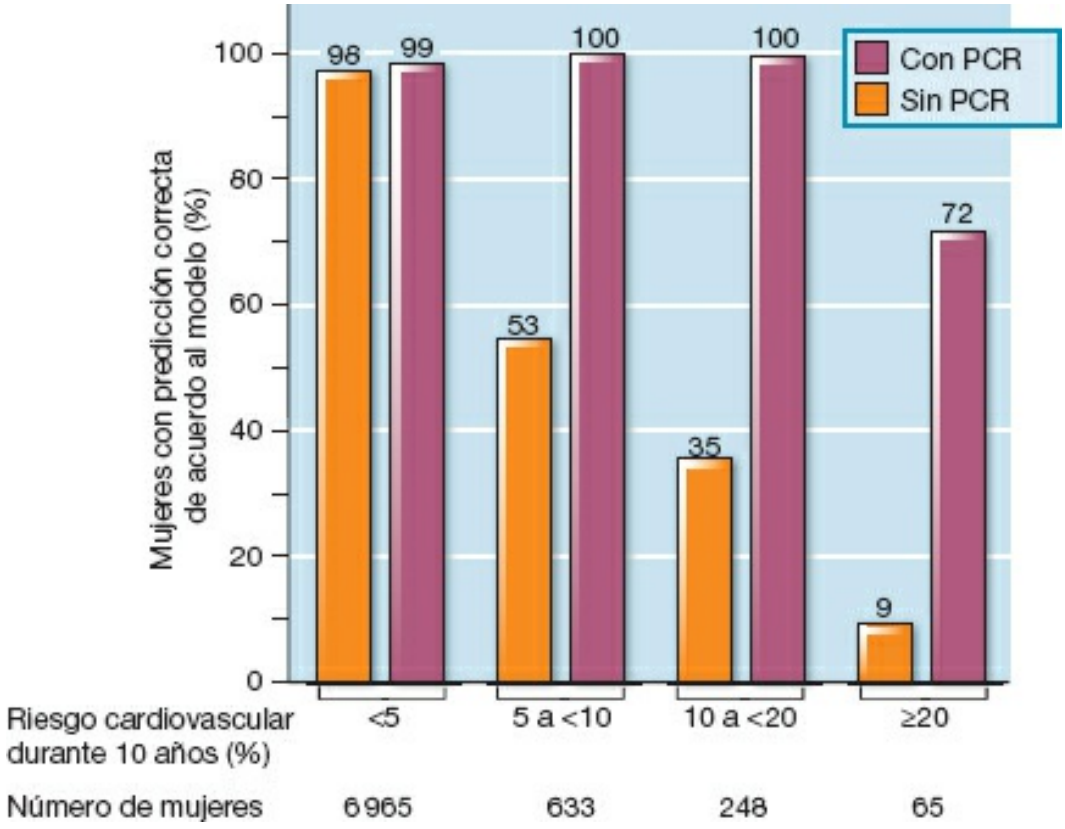


Figura 4-2 ■ Efecto de añadir un factor de riesgo nuevo a un modelo de predicción de riesgo. Comparación de modelos de predicción de riesgo para enfermedad cardiovascular (ECV) durante 10 años entre 7 911 mujeres sin diabetes, con y sin la proteína C reactiva (PCR) como un factor. La adición de la PCR en el modelo de riesgo mejoró la estratificación del riesgo de las mujeres, especialmente en la clasificación de mayor riesgo con el modelo sin la PCR. (Datos de Ridker PM, Buring JE, Rifal N et al. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women. JAMA 2007;297:611-619.)

Tabla 4-1

Ejemplo de una herramienta de predicción de riesgo: la herramienta de Evaluación del riesgo de cáncer de mama del NCI[a]

Factores de riesgo incluidos en el modelo[b]

1. ¿Qué edad tiene la mujer?
2. ¿Qué edad tenía en el momento de su primer período menstrual (menarquia)?
3. ¿Qué edad tenía en el momento del nacimiento de su primer hijo vivo?

booksmedicos.org

4. ¿Cuántas familiares de primer grado (madre, hermanas, hijas) han tenido cáncer de mama?
5. ¿Alguna vez le han realizado una biopsia mamaria?
5a. ¿A cuántas biopsias mamarias (positivas o negativas) se ha sometido?
5b. ¿Ha tenido al menos una biopsia mamaria con hiperplasia atípica?
6. ¿Cuál su raza/grupo étnico?
6a. ¿A qué subtipo racial/étnico pertenece?

[a]La herramienta de evaluación del riesgo no es para mujeres con antecedente de cáncer de mama, carcinoma ductal *in situ* o carcinoma lobulillar *in situ*.
[b]Una mujer (o su médico) elijen las respuestas para cada pregunta de un menú desplegable.
Disponible en: http://www.cancer.gov/bcrisktool/.

Predicción del riesgo en pacientes individuales y grupos

Las herramientas de predicción del riesgo se usan a menudo para predecir el futuro de las personas, con la esperanza de que cada persona conozca sus riesgos, una esperanza resumida por el término «medicina personalizada». Como ejemplo, la tabla 4-1 sintetiza la información utilizada en la herramienta de evaluación del riesgo de cáncer de mama del NCI. Una mujer o su médico introducen la información y la herramienta calcula el riesgo de desarrollar cáncer de mama a los 5 años y a lo largo de la vida (hasta los 90 años). Sin embargo, es mucho más difícil predecir lo que sucederá en un solo individuo que en un grupo de personas similares.

En primer lugar, debido a que las predicciones se expresan como probabilidades de episodios futuros, existe una incompatibilidad básica entre la incidencia de la enfermedad durante 5 años (p. ej., 15 %) en un grupo de personas y la probabilidad que un individuo del grupo desarrolle la enfermedad. Una sola persona puede o no desarrollar la enfermedad. (No se puede estar «algo embarazada»). Así que, en cierto sentido, el promedio del grupo siempre está mal para una persona porque los dos se expresan en términos diferentes, una probabilidad determinada por lo que le pasó a un grupo en el pasado frente a la predicción futura de la presencia o la ausencia de la enfermedad en una persona.

En segundo lugar, incluso la presencia de un factor de riesgo importante no necesariamente significa que una persona tenga la enfermedad. Como se destacó antes, muchos años de tabaquismo pueden aumentar el riesgo de cáncer de pulmón en un fumador unas 20 veces en comparación con las personas que no fuman. Aun así, el fumador tiene una probabilidad aproximada de 1 entre 10 de desarrollar cáncer de pulmón en los próximos 10 años. La mayoría de los factores de riesgo (y las herramientas de predicción

booksmedicos.org

del riesgo) para gran parte de las enfermedades son mucho más débiles que el riesgo de desarrollar cáncer de pulmón con el tabaquismo.

EVALUACIÓN DE LAS HERRAMIENTAS DE PREDICCIÓN DE RIESGO

Se puede determinar si una herramienta de predicción de riesgo en particular funciona bien formulando dos preguntas: *i)* ¿Cuál es la precisión de la herramienta para predecir la *proporción* de personas de diferentes grupos que desarrollarán la enfermedad? (**calibración**), y *ii)* ¿Cuál es la precisión con la que identifica a los *individuos* que desarrollarán o no la enfermedad? (**discriminación**). Para responder a estas preguntas, la herramienta debe ser probada en un gran grupo de personas que han sido vigiladas durante varios años (algunas veces, por décadas), conociendo el resultado de la enfermedad en cada persona del grupo.

Calibración

La calibración, que determina el grado en que la herramienta de predicción predice de forma correcta la proporción de un grupo que desarrollará la enfermedad, es conceptual y operativamente simple. Se mide comparando el número predicho o estimado (E) de personas, de acuerdo con la herramienta de predicción, que desarrollarán la enfermedad en un grupo contra el número observado (O) de personas que desarrollaron la enfermedad. Una proporción E/O cercana a 1,0 significa que la herramienta de riesgo está bien calibrada, predice una proporción de personas que se acerca mucho a la proporción real que desarrolla la enfermedad. Las evaluaciones de la herramienta de evaluación del riesgo de cáncer de mama del NCI predicen con gran precisión la proporción de mujeres de un grupo que desarrollarán cáncer de mama en los siguientes 5 años, con una proporción E/O cercana a 1,0.

Discriminación

Es difícil discriminar entre los individuos de un grupo que desarrollarán o no la enfermedad, incluso para las herramientas de predicción de riesgo bien calibradas. El método que se usa más frecuentemente para medir la precisión de la discriminación es calcular la **concordancia estadística** (a menudo abreviada como **estadística c**). En parejas de individuos seleccionados al azar, uno de los cuales desarrolló la enfermedad de interés mientras que el

booksmedicos.org

otro no, estima la frecuencia con la que una puntuación fue mayor para predecir el riesgo en un individuo que desarrolló la enfermedad. Si la herramienta de predicción de riesgo no mejora la predicción en absoluto, la estimación resultante sería igual que lanzar una moneda al aire y la estadística c sería de 0,50. Si la herramienta funcionara perfectamente, de manera que en cada pareja de individuos la persona enferma tiene una puntuación más alta que la persona no enferma, la estadística c sería 1,0. Un estudio evaluó la discriminación de la herramienta de riesgo para cáncer de mama del NCI con un estadístico c de 0,58 (9). Está claro que no es un estadístico c elevado, pero se debe considerar que es difícil comprender el significado clínico de los valores entre 0,5 y 1,0.

El método más claro (aunque el más raro) para comprender si un modelo de predicción del riesgo es bueno para discriminar es comparar visualmente las predicciones individuales con los resultados observados para todos los individuos del estudio. La figura 4-3 A muestra una discriminación perfecta por una herramienta hipotética de predicción del riesgo; la herramienta separa por completo a las personas destinadas a desarrollar la enfermedad de aquellas que no lo están. La figura 4-3 B ilustra la habilidad de la herramienta de predicción de riesgo de cáncer de mama del NCI para discriminar entre las mujeres que después desarrollaron o no cáncer de mama durante 5 años y muestra lo que significa un estadístico c de 0,58. Aunque las puntuaciones promedio del riesgo son algo mayores para las mujeres que desarrollaron cáncer de mama y su curva en la gráfica está ligeramente hacia la derecha para aquellas que no lo desarrollaron, las puntuaciones individuales de predicción del riesgo en todos los grupos se solapan sustancialmente; no existe un sitio en el eje *x* de riesgo que separe a las mujeres en grupos que desarrollarán o no cáncer de mama. Esto es así, aunque el modelo tuvo una calibración muy buena.

Sensibilidad y especificidad de una herramienta de predicción del riesgo

Otra manera de evaluar la habilidad de una herramienta de predicción del riesgo para discriminar entre quienes desarrollarán o no la enfermedad es determinar su sensibilidad y especificidad (se discutirá el tema a fondo en los capítulos 8 y 10). La **sensibilidad** de una herramienta de predicción del riesgo es la capacidad para identificar a los individuos destinados a desarrollar la enfermedad; se expresa como el porcentaje de personas que desarrollarán la enfermedad que la herramienta logra identificar de forma

booksmedicos.org

correcta. La **especificidad** de una herramienta es la capacidad de identificar a los individuos que no desarrollarán la enfermedad, expresada como el porcentaje de personas que no desarrollarán la enfermedad que la herramienta logra identificar correctamente. En relación con la figura 4-3, se eligió un riesgo a 5 años de 1,67 % como el punto de corte entre riesgo «bajo» y «alto». Usando este punto de corte, se estimó una sensibilidad del 44 % (el 44 % de las mujeres que desarrollaron cáncer de mama tuvieron una puntuación de riesgo ≥ 1,67 %) y una especificidad del 66 % (el 66 % de las mujeres que no desarrollaron cáncer de mama tuvieron una puntuación de riesgo < 1,67 %). En otras palabras, la herramienta de predicción de riesgo pasó por alto a más de la mitad de las mujeres que desarrollaron cáncer de mama en 5 años y asignó a un tercio de las mujeres no destinadas a desarrollar la enfermedad en el grupo de alto riesgo. Los estudios sobre las herramientas de predicción que incluyen información sobre la sensibilidad y especificidad también muestran con frecuencia la **curva característica operativa del receptor (ROC),** un método de análisis que combina los resultados de sensibilidad y especificidad y que puede ser utilizado para comparar diferentes herramientas. Las curvas ROC se comentan con mayor detalle en el capítulo 8.

booksmedicos.org

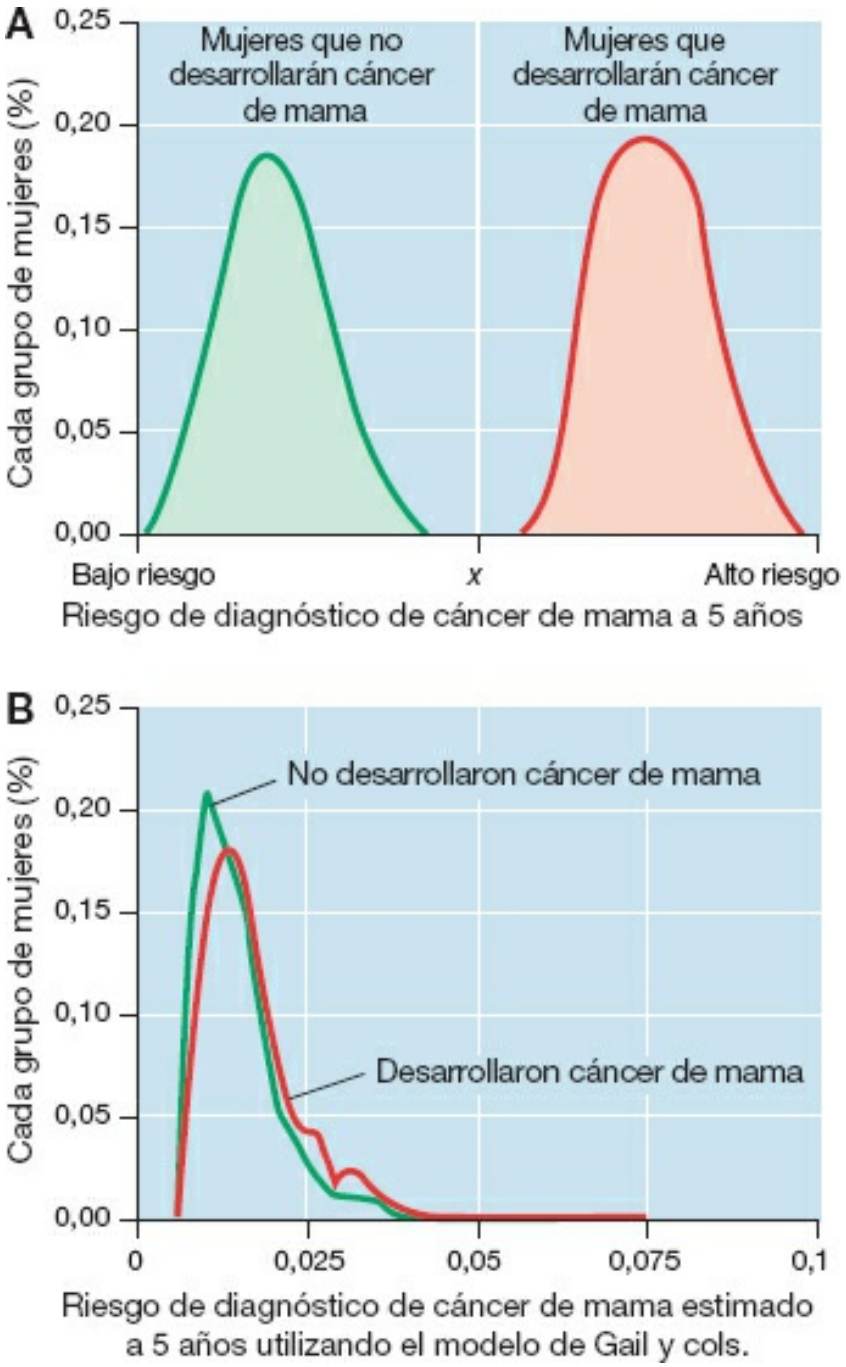


Figura 4-3 ■ A. La habilidad de una herramienta de predicción del riesgo hipotéticamente perfecta para discriminar entre las mujeres que desarrollaron o no cáncer de mama. El grupo de la izquierda tiene puntuaciones de riesgo bajo y no desarrollaron cáncer de mama, mientras que el grupo de la derecha tiene puntuaciones más altas y desarrollaron cáncer de mama. Estos dos grupos no se solapan y el estadístico c sería de 1,0. (Redibujado con autorización de Elmore JA, Fletcher SW. The risk of cancer risk prediction: «What is my risk of getting breast cancer?» J Natl Cancer Inst 2006;98:1673-1675.) **B. La habilidad de una herramienta de predicción del riesgo real para discriminar entre las mujeres que desarrollaron o no cáncer de mama en 5 años.** Las puntuaciones de riesgo entre los dos grupos se solapan sustancialmente, no existe un sitio en el eje *x* de riesgo que separe a las mujeres en grupos que desarrollarán o no cáncer de mama. (Redibujado con autorización de

booksmedicos.org

Rockhill Levine B.)

Estratificación del riesgo

Como se mencionó antes y como se muestra en la figura 4-2, la estratificación del riesgo se puede utilizar para evaluar si una herramienta de predicción del riesgo trabaja bien y para determinar si la adición de un factor de riesgo nuevo mejora la habilidad de la herramienta para clasificar de forma correcta a las personas en grupos de riesgo clínicamente significativos. Una buena estratificación del riesgo mejora la calibración de la herramienta. Es probable que la estratificación del riesgo no afecte dramáticamente a la capacidad de discriminación de la herramienta. Por ejemplo, al examinar la figura 4-2, la herramienta de riesgo que incluye a la PCR asignó correctamente al 99 % de 6 965 mujeres en la clasificación más baja de riesgo (< 5 % de episodios de ECV durante 10 años). El estudio encontró que los episodios de ECV ocurrieron en 101 mujeres (1,4 %) asignadas a la clasificación más baja de riesgo, un resultado consistente con < 5 %. Sin embargo, debido a que la gran mayoría de las mujeres (88 %) fueron asignadas a la clasificación más baja de riesgo, más mujeres en este grupo desarrollaron ECV (101) que en todos los demás grupos de riesgo combinados (97). Este resultado, similar a lo ocurrido con la herramienta de predicción de riesgo de cáncer de mama (fig. 4-3 B) es un incidente habitual y frustrante con la predicción del riesgo.

¿Por qué las herramientas de predicción del riesgo no discriminan bien entre los individuos?

¿Por qué una herramienta de predicción del riesgo que predice tan bien la proporción de un grupo de personas que desarrollarán la enfermedad no es buena para discriminar entre aquellos que no desarrollarán la enfermedad? Un problema importante es la fortaleza (o más correctamente, la debilidad) de la herramienta de predicción. La discriminación necesita un factor de riesgo muy fuerte (o la combinación de factores de riesgo) para separar a las personas que desarrollarán o no la enfermedad dentro de un grupo, incluso con un éxito moderado. Si las que desarrollarán la enfermedad sólo tienen una probabilidad dos o tres veces mayor, o incluso cinco, las herramientas de predicción de riesgo no discriminarán bien. Deben tener muchas más probabilidades (algunos autores sugieren que al menos 200 veces [10]) de desarrollar la enfermedad. Pocas reglas de predicción del riesgo son potentes.

booksmedicos.org

Otro problema es que para la mayoría de las enfermedades crónicas los factores de riesgo se encuentran ampliamente distribuidos en la población. Por lo tanto, incluso las personas con bajo riesgo pueden desarrollar la enfermedad. La figura 4-3 B muestra que algunas mujeres con la puntuación de riesgo más baja desarrollaron cáncer de mama. De hecho, en números absolutos, más mujeres con puntaciones bajas desarrollaron cáncer de mama que aquellas con puntuaciones altas porque (menos mal) en la mayoría de los grupos de mujeres existen relativamente pocas mujeres con puntuaciones altas.

En resumen, los modelos de predicción del riesgo son una forma importante de combinar factores de riesgo individuales para lograr una mejor estratificación de las personas en grupos de riesgo graduados, como se ilustra en la figura 4-2. Para un paciente individual y su médico es útil comprender en qué grupo de riesgo se asigna al paciente un modelo de riesgo bien construido, pero también deben entenderse las limitaciones de esta asignación. Además, es importante tener en cuenta el hecho contradictorio que, para la mayoría de las enfermedades, la mayoría de las personas destinadas a desarrollar la enfermedad no tienen un alto riesgo.

USOS CLÍNICOS DE LOS FACTORES DE RIESGO Y LAS HERRAMIENTAS DE PREDICCIÓN DEL RIESGO

Factores de riesgo y probabilidad preprueba de enfermedad en las pruebas diagnósticas

El conocimiento del riesgo puede utilizarse en el proceso diagnóstico, puesto que la presencia de un factor de riesgo aumenta la probabilidad de enfermedad. Sin embargo, la mayoría de los factores de riesgo (e incluso las herramientas de predicción de riesgo) tienen una habilidad limitada para predecir la enfermedad en pacientes sintomáticos porque no suelen ser unos predictores de enfermedad tan potentes como los hallazgos clínicos de una enfermedad precoz. Como estableció Geoffrey Rose (11):

> A menudo, el mejor predictor de futuras enfermedades graves es la presencia de una enfermedad leve. Una función ventilatoria reducida es, en la actualidad, el mejor predictor de su futura tasa de empeoramiento. Una elevada presión arterial actual es el mejor predictor de una futura tasa de aumento. Una cardiopatía coronaria temprana es, de todos los factores de riesgo convencionales, el mejor para predecir una futura enfermedad mortal.

booksmedicos.org

A modo de ejemplo de la afirmación de Rose, aunque la edad, el género masculino, el tabaquismo, la hipertensión, la hiperlipidemia y la diabetes son predictores importantes de enfermedad coronaria futura, son mucho menos importantes cuando se evalúa a un paciente que se presenta en urgencias con dolor torácico (12). Los aspectos concretos de una situación clínica, como la presencia y el tipo de dolor torácico y los resultados de un electrocardiograma, son los primeros pasos más poderosos para determinar si un paciente está experimentando un infarto agudo de miocardio (13).

La ausencia de un factor de riesgo muy importante puede ayudar a descartar la enfermedad. Así pues, es razonable considerar el mesotelioma en el diagnóstico diferencial de una masa pleural en un paciente que ha trabajado expuesto al amianto (asbesto). Sin embargo, el diagnóstico de mesotelioma es mucho menos probable en un paciente que nunca ha estado en contacto con este material.

Uso de los factores de riesgo para elegir un tratamiento

Los factores de riesgo se han utilizado durante largo tiempo para elegir (y desarrollar) tratamientos. Los pacientes con ECV, que también tienen lípidos elevados, son tratados con estatinas u otros fármacos para disminuir los lípidos. Los tratamientos específicos para la hiperlipidemia y la hipertensión son altamente efectivos para los pacientes diabéticos con estos trastornos. En oncología se han desarrollado tratamientos «dirigidos» para ciertos tipos de cáncer.

Ejemplo

El receptor HER2 es un receptor del factor de crecimiento epidérmico que tiene una mayor expresión en un 20 % de los cánceres de mama. El cáncer de mama positivo a HER2 se asocia con un pronóstico más adverso que los cánceres negativos a HER2. Se desarrolló un anticuerpo monoclonal dirigido contra la proteína HER2 y se probó en varios ensayos clínicos donde participaron mujeres con cáncer de mama positivo a HER2. Las mujeres que recibieron tanto el tratamiento convencional como el tratamiento con el anticuerpo monoclonal tuvieron la mitad de los episodios de cáncer de mama en comparación con quienes recibieron sólo el tratamiento convencional (14,15). Actualmente, el tamizaje de la proteína HER2 es un procedimiento estándar que ayuda a determinar el tratamiento óptimo para el cáncer de mama.

Estratificación del riesgo en los programas de tamizaje

booksmedicos.org

El conocimiento de los factores de riesgo en ocasiones puede utilizarse para mejorar la eficacia de los programas de tamizaje mediante la selección de subgrupos de pacientes con un riesgo sustancialmente mayor. A pesar de que el riesgo de cáncer de mama debido a mutaciones genéticas nocivas es muy reducido en la población general, dicho riesgo es mucho más elevado en aquellas mujeres cuyos familiares cercanos presentaron la enfermedad a una edad relativamente precoz; el tamizaje de mutaciones genéticas mediante análisis de sangre suele reservarse a las mujeres cuyos antecedentes familiares indican una situación de riesgo elevado. De igual modo, se recomienda que la población general se someta a una detección sistemática de cáncer colorrectal a partir de los 50 años. Sin embargo, las personas con un familiar de primer grado con antecedentes de cáncer colorrectal presentan un mayor riesgo de sufrir la enfermedad, y varios grupos de expertos sugieren que dichas personas deberían someterse al tamizaje o detección precoz a partir de los 40 años.

Eliminar factores de riesgo para prevenir la enfermedad

Si un factor de riesgo constituye también una causa de enfermedad, su eliminación puede evitar dicha enfermedad. La prevención puede ocurrir independientemente de que se conozca o no el mecanismo por el que se produce esta. Algunos de los logros clásicos de la historia de la epidemiología ilustran este punto. Antes de que se identificaran las bacterias, John Snow observó en 1854 que la tasa de cólera aumentaba entre las personas que bebían el agua suministrada por una determinada empresa, y la epidemia remitió tras suspender dicho suministro. En el proceso, estableció que el cólera se propagaba a través de los suministros de agua contaminada. En épocas más recientes se utiliza el mismo método para investigar los brotes de enfermedades transmitidas por alimentos, para identificar su origen y adoptar las medidas pertinentes para detenerlo. Hoy día, la causa biológica también se determina rápidamente y ayuda a identificar la fuente de la epidemia. El concepto de causa y su relación con la prevención se tratarán en el capítulo 12.

Preguntas de revisión

booksmedicos.org

Seleccione la mejor respuesta para las preguntas 4-1 a 4-10.

4-1. A mediados del s. xx, los cirujanos torácicos de Inglaterra tenían la impresión de estar operando más hombres con cáncer de pulmón, la mayoría de los cuales fumaban. ¿Cómo podría ser errónea la impresión de los cirujanos de que fumar era un factor de riesgo para desarrollar cáncer de pulmón?

A. Fumar se había vuelto tan común que más hombres tendrían el antecedente de tabaquismo, independientemente de si se sometieron a operaciones para el cáncer de pulmón.
B. El cáncer de pulmón es infrecuente, incluso entre los fumadores.
C. El tabaquismo confiere un riesgo bajo de cáncer de pulmón.
D. Existen otros factores de riesgo para el cáncer de pulmón.

4-2. Los factores de riesgo pueden desarrollarse fácilmente:

A. Cuando la exposición al factor de riesgo se produce mucho antes que la enfermedad.
B. Cuando la exposición al factor de riesgo se asocia con una nueva enfermedad.
C. Cuando el factor de riesgo es un marcador en vez de la causa de la enfermedad.

4-3. Los modelos de predicción del riesgo son útiles para:

A. Predecir el inicio de la enfermedad.
B. Diagnosticar la enfermedad.
C. Predecir el pronóstico.
D. Todas las anteriores.

4-4. La figura 4-2 muestra:

A. Que el modelo de riesgo que incorpora los resultados de la PCR asigna demasiadas mujeres al grupo de riesgo intermedio.
B. Que el modelo de riesgo que incorpora los resultados de la PCR predice mejor a las mujeres que desarrollarán ECV que el modelo de riesgo sin los resultados de la PCR.
C. El número de mujeres que desarrollarán ECV en 10 años es

booksmedicos.org

probablemente mayor en el grupo con un riesgo < 5 %.

4-5. Un modelo de riesgo para cáncer de colon estima que uno de sus pacientes tiene una probabilidad del 2 % de desarrollar cáncer colorrectal en los siguientes 5 años. Al explicar este hallazgo a su paciente, ¿cuál de las siguientes afirmaciones es la más correcta?

A. Ya que el cáncer colorrectal es el segundo cáncer más habitual en los hombres, excluyendo el cáncer de piel, debe estar preocupado por este hallazgo.

B. El modelo muestra que su paciente no desarrollará cáncer colorrectal en los siguientes 5 años.

C. El modelo muestra que su paciente pertenece al grupo donde un número muy reducido de personas desarrollarán cáncer colorrectal en los siguientes 5 años.

4-6. En general, las herramientas de predicción de riesgo son mejores para:

A. Predecir la enfermedad futura en un paciente dado.

B. Predecir la enfermedad futura en un grupo de pacientes.

C. Predecir los individuos que desarrollarán o no la enfermedad.

4-7. Cuando un factor de riesgo es un marcador de enfermedad futura:

A. El factor de riesgo ayuda a identificar a las personas que tienen mayor riesgo de desarrollar la enfermedad.

B. Eliminar el factor de riesgo puede ayudar a prevenir la enfermedad.

C. El factor de riesgo no está confundiendo una verdadera relación causal.

4-8. Un factor de riesgo generalmente tiene menos utilidad para:

A. El proceso de estratificación del riesgo.

B. Diagnosticar un malestar del paciente.

C. Prevenir la enfermedad.

4-9. La figura 4-3 B muestra que:

A. El modelo de riesgo está bien calibrado.

B. El modelo de riesgo funciona bien para estratificar a las mujeres en

booksmedicos.org

diferentes grupos de riesgo.

C. La mayoría de las mujeres que desarrollarán cáncer de mama a lo largo de 5 años tienen un mayor riesgo.

D. El modelo de riesgo no discrimina muy bien.

4-10. Es difícil que los modelos de riesgo determinen a los individuos que desarrollarán o no la enfermedad debido a las razones siguientes, *excepto:*

A. La combinación de factores de riesgo no está fuertemente relacionada con la enfermedad.

B. Los factores de riesgo son comunes en toda la población.

C. El modelo está bien calibrado.

D. La mayoría de las personas destinadas a desarrollar la enfermedad no tienen alto riesgo.

Encontrará las respuestas en el apéndice A.

BIBLIOGRAFÍA

1. Vineis P, Alavanja M, Buffler P, et al. Tobacco and cancer: recent epidemiological evidence. *J Natl Cancer Inst* 2004;96:99–106.
2. Willinger M, Hoffman HJ, Wu KT, et al. Factors associated with the transition to nonprone sleep positions of infants in the United States: The National Infant Sleep Position Study. *JAMA* 1998;280:329–335.
3. Li DK, Petitti DB, Willinger M, et al. Infant sleeping position and the risk of sudden infant death syndrome in California, 1997–2000. *Am J Epidemiol* 2003;157:446–455.
4. Chen WY, Rosner B, Hankinson SE, et al. Moderate alcohol consumption during adult life, drinking patterns and breast cancer risk. *JAMA* 2011;306:1884–1890.
5. Humphrey LL, Rongwei F, Rogers K, et al. Homocysteine level and coronary heart disease incidence: a systematic review and meta-analysis. *Mayo Clin Proc* 2008;83:1203–1212.
6. Martí-Carvajal AJ, Solà I, Lathyris D, et al. Homocysteine lowering interventions for preventing cardiovascular events. *Cochrane Database Syst Rev* 2009;4:CD006612.
7. Buckley DI, Fu R, Freeman M, et al. C-reactive protein as a risk factor for coronary heart disease: a systematic review and meta-analyses for the U.S. Preventive Services Task Force. *Ann Intern Med* 2009;151:483–495.
8. Ridker PM, Hennekens CH, Buring JE, et al. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *N Engl J Med* 2000;342: 836–843.
9. Rockhill B, Speigelman D, Byrne C, et al. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst* 2001;93:358– 366.
10. Wald NJ, Hackshaw AK, Frost CD. When can a risk factor be used as a worthwhile screening test? *BMJ* 1999;319: 1562–1565.
11. Rose G. Sick individuals and sick populations. *Int J Epidemiol* 1985;14:32–38.
12. Han JH, Lindsell CJ, Storrow AB, et al. The role of cardiac risk factor burden in diagnosing acute coronary syndromes in the emergency department setting. *Ann Emerg Med* 2007;49: 145–152.

booksmedicos.org

13. Panju AA, Hemmelgarn BR, Guyatt GH, et al. Is this patient having a myocardial infarction? The rational clinical examination. *JAMA* 1998;280:1256–1263.
14. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 2005;353:1659–1672.
15. Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 2005;353:1673–184.

booksmedicos.org

Capítulo 5

Riesgo: exposición a la enfermedad

A partir del estudio de las características de las personas que más tarde desarrollarán enfermedad cardiovascular (ECV) y la comparación de características de aquellos que permanecen sin la enfermedad, es posible hacer un perfil de aquellas personas con un alto riesgo de desarrollar ECV muchos años antes de que se manifiesten los síntomas o signos… En el caso de las enfermedades no infecciosas, rara vez ha sido posible identificar a tales individuos altamente susceptibles años antes que desarrollen la enfermedad.

—Thomas Dawber y William Kannel
1961

PALABRAS CLAVE

Estudio observacional
Estudio de cohorte
Cohorte
Grupo expuesto
Grupo no expuesto
Estudio de incidencia
Estudio de cohorte prospectivo
Estudio de cohortes retrospectivo/histórico
Diseño de casos y cohortes
Medidas de efecto
Riesgo absoluto
Riesgo atribuible
Diferencia de riesgos
Riesgo relativo
Razón de riesgos
Riesgo atribuible poblacional
Fracción atribuible poblacional
Variable extraña

booksmedicos.org

Covariable
Medida cruda del efecto
Confusión
Desenlace intermedio
Variable de confusión/confusión
Control
Restricción
Pareamiento
Estratificación
Estandarización
Análisis multivariable o multivariado
Regresión logística
Riesgo proporcional de Cox
Factor de confusión no medido
Modificación del efecto
Interacción

ESTUDIOS DE RIESGO

Este capítulo describe cómo los investigadores obtienen la estimación del riesgo, observando la relación entre los posibles factores de riesgo y la incidencia subsecuente de la enfermedad. En él se describen los métodos utilizados para determinar el riesgo al seguir a los grupos en el futuro y también se discuten varias formas de comparar los riesgos que afectan a los individuos y a las poblaciones. En el capítulo 6 se describen los métodos para estudiar el riesgo al revisar los acontecimientos del pasado.

El método más potente para determinar si la exposición a un posible factor de riesgo da lugar a un mayor riesgo de la enfermedad consiste en realizar un experimento en el que el investigador determine quiénes están expuestos. Las personas que al inicio no sufren la enfermedad se dividen en dos grupos con la misma susceptibilidad a la enfermedad en cuestión. Uno de los grupos se expone al supuesto factor de riesgo, mientras que el otro no se somete a la exposición; por lo demás, el tratamiento de los grupos es el mismo. Posteriormente, cualquier diferencia observada en las tasas de enfermedad de los grupos podría atribuirse al factor de riesgo. Los experimentos se exponen en el capítulo 9.

Cuando no es posible realizar experimentos

Los efectos de la mayoría de los factores de riesgo en los humanos no pueden evaluarse mediante estudios experimentales. Considérense algunas

booksmedicos.org

de las preguntas relacionadas con el riesgo que nos inquietan en la actualidad: ¿Presentan las personas inactivas un mayor riesgo de sufrir enfermedad cardiovascular (ECV), siendo el resto de los factores idénticos? ¿Los teléfonos móviles provocan tumores cerebrales malignos? ¿Una alimentación con un elevado contenido de grasas aumenta el riesgo de sufrir cáncer de mama? Para responder a estas preguntas, habitualmente no puede realizarse un experimento. En primer lugar, no sería ético imponer posibles factores de riesgo a un grupo de personas sanas para someterlas a una investigación científica. En segundo lugar, la mayoría de ellas se negaría a que otras personas limitaran sus dietas y conductas durante largos períodos. Finalmente, el experimento tendría que durar muchos años, lo que resultaría dificultoso y caro. Por todo ello, suele ser necesario estudiar el riesgo con métodos menos molestos.

Los estudios clínicos en los que el investigador recoge los datos tras, sencillamente, observar los acontecimientos conforme tienen lugar, sin desempeñar un papel activo en lo que sucede, se denominan **estudios observacionales.** La mayoría de los estudios de riesgo son de este tipo, y pueden ser**estudios de cohortes,** que se describen en lo que queda de capítulo, o estudios que se describen en el capítulo 6.

Cohortes

Tal como se definió en el capítulo 2, el término **cohorte** se utiliza para describir un grupo de personas que tienen algo en común en el momento en que se agrupan y que son seguidas durante un período de tiempo para observar qué les sucede. En la tabla 5-1 se enumeran algunas de las utilidades de las cohortes en la investigación clínica. Independientemente de lo que tengan en común los miembros de la cohorte, si las observaciones están destinadas a proporcionar información sólida sobre el riesgo de la enfermedad, dichos miembros deben cumplir tres criterios.

Tabla 5-1

Cohortes y sus propósitos

Característica en común	Para evaluar el efecto de	Ejemplo
Edad	Edad	Esperanza de vida de las personas de 70 años

booksmedicos.org

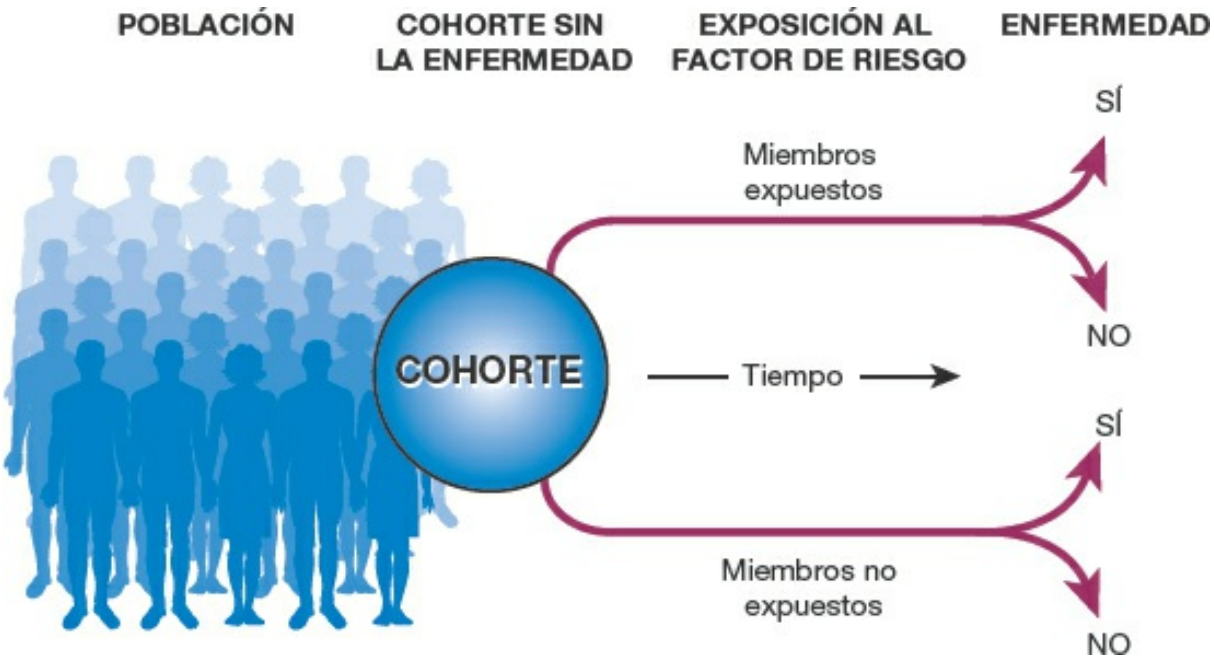
		(independientemente de la fecha de nacimiento)
Fecha de nacimiento	Un período concreto	Tasas de tuberculosis en personas nacidas en 1930
Exposición	Factor de riesgo	Cáncer de pulmón en personas que fuman
Enfermedad	Pronóstico	Tasa de supervivencia en pacientes con cáncer de mama
Intervención terapéutica	Tratamiento	Mejora de la supervivencia de pacientes con enfermedad de Hodgkin que reciben quimioterapia combinada
Intervención preventiva	Prevención	Reducción de la incidencia de neumonía tras recibir vacunación antineumocócica

1. No deben sufrir la enfermedad en cuestión en el momento de su agrupación.
2. Deben ser observados a lo largo de un período de tiempo importante dentro de la evolución natural de la enfermedad en cuestión, con el fin de que se disponga de tiempo suficiente para que el riesgo se exprese. Por ejemplo, si se desea averiguar si la radiación en el cuello durante la infancia da lugar a neoplasias tiroideas, un seguimiento de 5 años no sería una prueba razonable para confirmar la hipótesis, puesto que el período habitual entre la exposición a la radiación y el inicio de la enfermedad es considerablemente mayor.
3. Todos los miembros de la cohorte deben ser observados durante todo el período de seguimiento o se deben usar métodos para considerar la deserción. En la medida en que las personas abandonen el estudio y sus razones para ello estén relacionadas de alguna manera con el resultado, una cohorte incompleta aportará información que puede representar de forma errónea el estado real de la situación.

booksmedicos.org

Estudios de cohorte

La figura 5-1 muestra el diseño básico de un estudio de cohorte. Un grupo de personas (una cohorte), ninguna de las cuales ha presentado el episodio de interés, aunque todas podrían presentarlo. (Por ejemplo, en un estudio de los factores de riesgo del cáncer de endometrio, cada miembro de la cohorte debe contar con un útero sano.) Al ser incluidos en el estudio, los miembros de la cohorte se clasifican en función de aquellas características (posibles factores de riesgo) que podrían estar relacionadas con el resultado. Para cada posible factor de riesgo, los miembros se clasifican como **expuestos** (poseen el factor en cuestión, p. ej., hipertensión arterial) o **no expuestos.** Todos los miembros de la cohorte son observados durante un tiempo para detectar quiénes presentan el resultado (en el ejemplo, ECV), y se comparan las tasas de los resultados de interés en los grupos expuestos y no expuestos. Es en ese momento cuando se puede observar cómo los posibles factores de riesgo se relacionan con resultados de interés posteriores. Otros nombres para los estudios de cohorte son **estudios de incidencia,** denominación que destaca que los pacientes son seguidos a lo largo del tiempo; estudios prospectivos, que implica la dirección en que se realiza el seguimiento de los pacientes, es decir, hacia delante, y los estudios longitudinales, que se enfocan en la medida básica de nuevos episodios de una enfermedad a lo largo del tiempo.



booksmedicos.org

Figura 5-1 ■ Diseño de un estudio de cohorte sobre el riesgo. Las personas sin la enfermedad se clasifican en dos grupos: las que están expuestas al factor de riesgo y las que no lo están. Se realiza el seguimiento de ambos grupos a lo largo del tiempo para determinar qué proporción de miembros de cada grupo sufre la enfermedad.

A continuación se describe un estudio de cohorte clásico que ha realizado importantes aportaciones a nuestra comprensión de los factores de riesgo de la ECV y los métodos modernos para realizar estudios de cohorte.

Ejemplo

El estudio Framingham (1) se inició en 1949 para identificar factores asociados a un mayor riesgo de cardiopatía coronaria (CC). Se seleccionó una muestra representativa de 5 209 hombres y mujeres, con edades entre los 30 y los 59 años, de entre 10 000 personas de dicha edad que vivían en Framingham, una pequeña ciudad situada cerca de Boston (EE.UU.). De entre estas personas, 5 127 no sufrían CC cuando fueron examinadas por primera vez y, por lo tanto, presentaban un riesgo de presentarla. Se las examinó de forma bienal en busca de signos de enfermedad coronaria. El estudio duró 30 años y actualmente se continúa con el estudio Framingham Offspring (2). Demostró que el riesgo de sufrir CC se asociaba a una elevada presión arterial, un alto nivel de colesterol sérico, tabaquismo, intolerancia a la glucosa e hipertrofia ventricular izquierda. Se produjo una gran diferencia en cuanto al riesgo de CC entre las personas con todos los factores de riesgo mencionados y las que no tenían ninguno de ellos. La combinación de los factores de riesgo identificados en este estudio dieron origen a una de las herramientas de predicción de riesgo más utilizadas en la medicina clínica: la puntuación de riesgo de Framingham para ECV.

Estudios de cohorte prospectivos y de cohorte históricos

Los estudios de cohorte pueden realizarse de dos formas (fig. 5-2). Puede constituirse la cohorte en el presente y realizarse su seguimiento en el futuro (**estudio de cohorte prospectivo**), o puede identificarse a partir de historias clínicas antiguas y seguirse desde ese momento concreto del pasado hasta el presente (**estudio de cohorte histórico o estudio de cohorte retrospectivo**). El estudio Framingham descrito es un ejemplo de estudio de cohorte prospectivo. En la literatura médica cada vez aparecen más estudios de cohorte retrospectivos útiles, debido a la disponibilidad de grandes bases de datos clínicas electrónicas.

Estudios de cohorte prospectivos

booksmedicos.org

Los estudios de cohorte prospectivos pueden evaluar los supuestos factores de riesgo que generalmente no se registran en los expedientes médicos, incluyendo muchas de las conductas relacionadas con la salud, el nivel educativo y el estado socioeconómico, pues se ha encontrado que tienen efectos importantes sobre la salud. Cuando el estudio se planifica antes de recopilar los datos, los investigadores pueden asegurarse de registrar la información de los posibles factores de confusión. Por último, toda la información de un estudio de cohorte prospectivo puede ser recopilada de una forma estandarizada que disminuye el sesgo de medición.

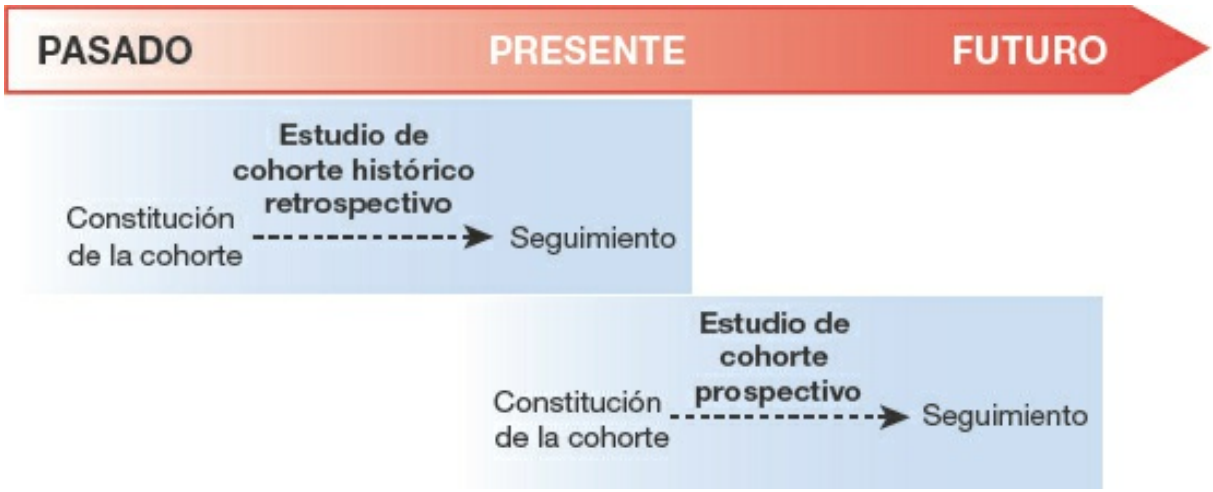


Figura 5-2 ■ Estudios de cohorte prospectivos y retrospectivos. En los estudios de cohorte prospectivos, las cohortes se constituyen en el presente y son seguidas en el futuro. Por el contrario, en los estudios de cohorte retrospectivos, las cohortes se constituyen retrocediendo al pasado, por ejemplo, a partir de historias clínicas, y su seguimiento se realiza hasta el presente.

Ejemplo

¿Cuánto tiempo libre dedicado a la actividad física es necesario para lograr beneficios para la salud? Varias directrices sugieren un mínimo de 30 min diarios durante 5 días a la semana, pero la mayoría de las personas no siguen esta recomendación. ¿Es posible que menos actividad física logre beneficios para la salud? Se realizó un estudio de cohorte prospectivo en el que más de 415 000 adultos respondieron un cuestionario estandarizado sobre su actividad física y fueron seguidos durante un promedio de 8 años (3). Los resultados mostraron que una mayor cantidad de actividad física se correlacionó con la reducción de la mortalidad por todas las causas y una mayor esperanza de vida en comparación con aquellos que no realizaron la actividad. Tan sólo 15 min de actividad diaria se correlacionaron con una disminución de la mortalidad del 14 % y un aumento en la expectativa de vida de 3 años, aun considerando el nivel educativo, el trabajo físico en el trabajo y otras condiciones médicas.

booksmedicos.org

Estudios de cohorte históricos utilizando bases de datos clínicas

Los estudios de cohorte históricos pueden aprovechar las bases de datos clínicas electrónicas y los registros de población que se utilizan principalmente para la atención del paciente o para averiguar la salud de la población. Las ventajas principales de estos estudios sobre los estudios clásicos de cohorte prospectivos es que tardan menos tiempo, son menos costosos y mucho más fáciles de realizar. Sin embargo, no se pueden estudiar los factores que no están registrados en las bases de datos electrónicas, por lo que en general no se pueden incluir en estos estudios factores como el estilo de vida, la educación y otros determinantes importantes de la salud del paciente. Además, la información de muchas bases de datos, sobre todo la información sobre la atención médica, no se recopila de una forma estandarizada, lo que conduce a la posibilidad de sesgo en los resultados. Las grandes bases de datos electrónicas son particularmente útiles para estudiar posibles factores de riesgo y desenlaces sobre la salud que es posible que se hayan registrado en las bases de datos clínicas de una forma un tanto estandarizada, como los diagnósticos y tratamientos.

Ejemplo

La incidencia del autismo aumentó bruscamente en la década de 1990, coincidiendo con una mayor administración de la vacuna triple vírica (sarampión, rubéola y parotiditis) a los niños pequeños. Un informe vinculó la vacuna triple vírica con el autismo. Debido a los casos en algunos niños, se extendió la alarma de que la vacuna (o el conservante que esta incorporaba, el timerosal) era la responsable del aumento de la incidencia del autismo. En algunos países descendió la tasa de administración de dicha vacuna entre los niños pequeños, lo que causó nuevos brotes e incluso muertes por sarampión. Debido a la gravedad de la situación, se realizaron varios estudios para evaluar la vacuna triple vírica como un posible factor de riesgo. En Dinamarca se llevó a cabo un estudio de cohorte retrospectivo en el que se incluyeron todos los niños (537 303) nacidos entre enero de 1991 y diciembre de 1998 (4). Los investigadores revisaron las historias clínicas de los niños en todo el país y determinaron que el 82 % recibió la vacuna (los médicos deben notificar la administración de vacunas al gobierno para poder cobrar); se diagnosticaron 316 niños de autismo, y otros 422 de sufrir trastornos del espectro autista. La frecuencia del autismo entre los niños que se habían vacunado fue similar (de hecho, ligeramente menor) a la de los que no la recibieron. Esta conclusión, junto con otros estudios, aportó pruebas sólidas en contra de la sugerencia de que la vacuna triple vírica provoca autismo. Después, el estudio original que causó la alerta fue investigado por fraude y conflicto de intereses, y se retractó en el 2010 en la revista *The Lancet* (5).

booksmedicos.org

Estudios de caso-cohorte

Otro método que utiliza bases de datos médicas informatizadas en estudios de cohorte es el **diseño de caso-cohorte.** Conceptualmente, es una modificación del diseño de cohorte retrospectivo que se beneficia de la capacidad para determinar la frecuencia de una afección concreta en un gran grupo de personas. En un estudio de caso-cohorte, todas las personas expuestas se incluyen en la cohorte, pero sólo una pequeña muestra aleatoria de personas no expuestas se incluyen en el estudio y se les sigue para determinar si presentan el desenlace de interés. Por razones de eficiencia, el grupo de personas no expuestas es «enriquecido» con las personas que posteriormente presentan el episodio de interés (es decir, que se convierten en casos). Entonces, los resultados se ajustan para que reflejen las fracciones de muestreo utilizadas para obtener la muestra. Este eficaz enfoque de un estudio de cohorte necesita que se determinen frecuencias de los episodios en todo el grupo de personas no expuestas, de ahí la necesidad de una gran base de datos médica informatizada.

Ejemplo

La mastectomía profiláctica ¿protege a las mujeres con un mayor riesgo de sufrir cáncer de mama? Para evaluar esta cuestión se llevó a cabo un estudio de caso-cohorte en seis centros sanitarios, todos ellos con bases de datos informatizadas que contenían los diagnósticos y los procedimientos quirúrgicos a que se habían sometido sus miembros. Los investigadores identificaron a todas las 276 mujeres a las que se practicó una mastectomía profiláctica bilateral a lo largo de cierto número de años, y realizaron su control de seguimiento durante un tiempo para determinar si presentaban cáncer de mama. Para el grupo de comparación, los investigadores seleccionaron de forma aleatoria a un grupo similar de mujeres que no se habían sometido a la intervención, y enriquecieron la muestra con mujeres que posteriormente presentaron cáncer de mama. El «enriquecimiento» se logró averiguando quién, de entre las 666 800 mujeres elegibles, presentó cáncer de mama (es decir, mediante la búsqueda en la base de datos informatizada). Los investigadores seleccionaron aleatoriamente un 1 % de las mujeres del grupo de comparación de una determinada edad que sufrieron cáncer de mama[†], pero sólo un 0,01 % de las mujeres no lo presentaron. Durante el estudio se realizaron los ajustes para las fracciones de muestreo. Los resultados señalaron que la mastectomía profiláctica bilateral se asociaba a una reducción del 99 % de los casos de cáncer de mama en las mujeres en situación de riesgo elevado (6).

Ventajas e inconvenientes de los estudios de cohorte

booksmedicos.org

Cuando no es posible llevar a cabo un experimento, los estudios de cohorte sobre el riesgo, independientemente del tipo, son los mejores sustitutos disponibles de un experimento verdadero siempre que estén bien conducidos. Siguen la misma lógica que un ensayo clínico y permiten medir la exposición a un posible factor de riesgo, evitando cualquier posibilidad de sesgo, que podría presentarse si se determinara la exposición después de conocer el desenlace. La desventaja científica más importante de los estudios de cohorte (de hecho, de todos los estudios observacionales) es que están sujetos a muchos más sesgos potenciales que los experimentos. Es posible que las personas expuestas a cierto factor de riesgo en el curso natural de los acontecimientos difieran de muchas maneras de un grupo comparativo de personas no expuestas al factor. Si algunas de estas diferencias también se relacionan con la enfermedad de interés, pueden confundir cualquier asociación observada entre el presunto factor de riesgo y la enfermedad.

[†]Estrictamente hablando, el estudio era una modificación de un diseño de caso-cohorte estándar que incluiría todos los casos, no sólo el 1 %, de 26 800 cánceres de mama que se convirtió en el grupo de comparación de las mujeres no sometidas a la mastectomía profiláctica. Sin embargo, debido a que el cáncer de mama ocurre con frecuencia basta una muestra aleatoria del grupo.

Tabla 5-2

Ventajas e inconvenientes de los estudios de cohorte

Ventajas	**Inconvenientes**
Todos los tipos de estudio de cohorte	
El único modo de establecer la incidencia (es decir, el riesgo absoluto) directamente	Susceptibles de sesgos de confusión y otros
Siguen la misma lógica que la pregunta clínica: si las personas están expuestas, ¿contraen la enfermedad?	
Puede obtenerse la exposición sin el sesgo que podría producirse de conocerse el resultado de antemano antes de documentar la exposición	

booksmedicos.org

Pueden evaluar la relación entre la exposición y muchas enfermedades	
Estudios de cohorte prospectivos	
Se puede estudiar una amplia gama de posibles factores de riesgo	Ineficaces, debido a que es necesario incluir muchos más pacientes de los que presentan el episodio de interés; por lo tanto, no pueden utilizarse para enfermedades poco frecuentes
Se pueden recopilar datos demográficos y sobre el estilo de vida que no están disponibles en la mayoría de los registros médicos	Son caros debido a los recursos necesarios para estudiar un gran número de personas a lo largo del tiempo
Se pueden desarrollar formas estandarizadas para medir la exposición y el grado de exposición a los factores de riesgo	No se dispone de los resultados durante un largo período de tiempo
	Evalúan la relación entre la enfermedad y la exposición con sólo un número relativamente reducido de factores (es decir, los registrados al inicio del estudio)
Estudios de cohorte retrospectivos (históricos)	
Son más eficientes que los estudios de cohorte prospectivos porque los datos se han recopilado para otro propósito (es decir, durante la atención del paciente o para un registro)	Se puede estudiar una menor gama de posibles factores de riesgo que en los estudios de cohorte prospectivos
Más barato que los estudios de cohorte prospectivos porque no es necesario contar con recursos para seguir a mucha gente a través del tiempo	No se pueden examinar las características del paciente que no están disponibles en el conjunto de datos que se está utilizando

booksmedicos.org

Más rápidos que los estudios de cohorte prospectivos porque los resultados del paciente ya se han presentado	La medición de la exposición y el grado de exposición pueden no ser estandarizadas
Estudios de caso-cohorte	
Poseen todas las ventajas de los estudios de cohorte retrospectivos	Posee todas las desventajas de los estudios de cohorte retrospectivos
Son aún más eficientes que los estudios de cohorte retrospectivos porque solamente se analiza una muestra del grupo no expuesto	Dificultad para los lectores para entender los procesos de ponderación utilizados en el análisis

En la tabla 5-2 se resumen los usos, las fortalezas y las limitaciones de los diferentes tipos de estudios de cohorte. Muchas de las ventajas e inconvenientes se aplican sin importar el tipo. Sin embargo, las posibles dificultades sobre la calidad de los datos son diferentes entre los tres. En los estudios prospectivos se puede recopilar la información específicamente para los propósitos del estudio y con plena anticipación de lo que se necesita. Por lo tanto, es posible evitar los sesgos de medición y algunos de los factores de confusión que podrían debilitar la exactitud de los resultados. Sin embargo, en general se recoge la información de las cohortes históricas para otros fines, a menudo como parte de los registros médicos para la atención del paciente. Exceptuando las preguntas cuidadosamente seleccionadas, como la relación entre la vacunación y el autismo, es probable que los datos de los estudios de cohorte históricos no tengan la calidad suficiente para la investigación rigurosa.

Los estudios de cohorte prospectivos también pueden recopilar datos sobre el estilo de vida y otras características que pueden influir en los resultados, y pueden hacerlo de una manera estandarizada. Muchas de estas características no suelen estar disponibles en los estudios retrospectivos, y de caso-cohorte, y las que suelen estar disponibles no se han recopilado de manera estandarizada.

El principal inconveniente de los estudios de cohorte prospectivos es que cuando el resultado es infrecuente, como suele suceder en los estudios de riesgo, el estudio debe incluir un gran número de personas que deberán permanecer en observación durante un largo período antes de que se

booksmedicos.org

obtengan resultados. El hecho de tener que medir la exposición en muchas personas y vigilarlas durante años es poco eficiente cuando muy pocas finalmente desarrollarán la enfermedad. Por ejemplo, el estudio Framingham sobre ECV (la causa de muerte más frecuente en EE.UU.) fue el estudio de mayor tamaño de los de su clase cuando se inició. A pesar de todo, fue necesario realizar un seguimiento a más de 5 000 personas durante varios años antes de que pudieran publicarse las primeras conclusiones preliminares. Durante los primeros 8 años, sólo el 5 % de la población sufrió un episodio coronario. Los estudios retrospectivos y de caso-cohorte pueden tratar con el problema del tiempo, pero con frecuencia sacrifican el acceso a información importante y estandarizada.

Otro problema de los estudios de cohorte prospectivos deriva del hecho de que las personas incluidas en el estudio suelen «vivir en libertad» y no bajo el control de los investigadores. Para seguirles el rastro es necesario invertir una gran cantidad de esfuerzo y dinero, lo que hace que estos estudios sean caros, con un costo habitual de muchos millones de dólares, a veces cientos de millones.

Debido al tiempo y al dinero que precisan los estudios de cohorte, este método no puede utilizarse para todas las preguntas clínicas sobre riesgo, una de las razones principales de los esfuerzos para encontrar formas más eficientes, aunque fiables, de evaluar el riesgo, como los diseños retrospectivos y de casos y controles. Otro método, los estudios de casos y controles, se expone en el capítulo 6.

FORMAS DE EXPRESAR Y COMPARAR EL RIESGO

La expresión básica del riesgo es la incidencia, definida en el capítulo 2 como el número de nuevos casos de enfermedad que surgen durante un período concreto en una población definida que inicialmente no sufre la afección. En los estudios de cohorte se compara la incidencia de la enfermedad en dos o más grupos que difieren en cuanto a la exposición a un posible factor de riesgo. Para comparar los riesgos, suelen utilizarse varias medidas de la asociación entre la exposición y la enfermedad, conocidas como **medidas de efecto.** Dichas medidas representan diferentes conceptos de riesgo y se utilizan para propósitos diferentes. En la tabla 5-3 se exponen cuatro medidas de efecto junto con el riesgo absoluto y, en la tabla 5-4, se demuestra su utilización con un ejemplo sobre el riesgo de

booksmedicos.org

cáncer de pulmón entre fumadores y no fumadores.

Riesgo absoluto

El **riesgo absoluto** es la probabilidad de observar un episodio en una población objeto de estudio. Su valor es el mismo que el de la incidencia, y ambos términos se utilizan a menudo de forma indistinta. El riesgo absoluto es la mejor forma de que los médicos y los pacientes entiendan el modo en que los factores de riesgo pueden afectar a sus vidas. Así, como muestra la tabla 5-4, aunque el tabaquismo aumenta considerablemente las probabilidades de morir por un cáncer de pulmón, entre los fumadores el riesgo absoluto de morir por esta causa cada año en la población estudiada fue de 341,3 por cada 100 000 (3 a 4 muertes por cáncer de pulmón por cada 1 000 fumadores por año).

Tabla 5-3

Medidas de efecto

Expresión	Pregunta	Definición[a]
Riesgo absoluto	¿Cuál es la incidencia de la enfermedad en un grupo que inicialmente no sufre la afección?	$I = \frac{\#\text{ nuevos casos a lo largo de un período de tiempo dado}}{\#\text{ personas en el grupo}}$
Riesgo atribuible (diferencia de riesgo)	¿Cuál es la incidencia de enfermedad atribuible a la exposición?	$RA = I_E - I_{\bar{E}}$
Riesgo relativo (razón de riesgos)	¿Cuántas veces es más probable que las personas expuestas presenten la enfermedad, en relación con las personas no expuestas?	$RR = \frac{I_E}{I_{\bar{E}}}$
Riesgo atribuible poblacional	¿Cuál es la incidencia de enfermedad en una población, asociada a la prevalencia de un factor de riesgo?	$RA_P = RA \times P$
Fracción atribuible poblacional	¿Qué fracción de la enfermedad en una población es atribuible a la exposición a un factor de riesgo?	$FA_P = \frac{RA_P}{I_T}$

[a]Donde I_E = incidencia en personas expuestas; $I_{\bar{E}}$ = incidencia en las personas no expuestas; P = prevalencia de la exposición a un factor de riesgo; I_T = incidencia total de la enfermedad en una población.

Tabla 5-4

Cálculo de las medidas del efecto: consumo de cigarrillos y muerte por cáncer de pulmón[a]

Riesgos simples	
Tasa de mortalidad (riesgo absoluto o incidencia) por cáncer de pulmón en fumadores	341,3/100 000/año

booksmedicos.org

Tasa de mortalidad (riesgo absoluto o incidencia) por cáncer en no fumadores de cigarrillos	14,7/100 000/año
Prevalencia del consumo de cigarrillos	32,1 %
Tasa de mortalidad por cáncer de pulmón en la población	119,4/100 000/año
Riesgos comparados	
Riesgo atribuible	= 341,3/100 000/año – 14,7/100 000/año = 326,6/100 000/año
Riesgo relativo	= 341,3/100 000/año ÷ 14,7/100 000/año = 23,2
Riesgo atribuible poblacional	326,6/100 000/año × 0,321 = 104,8/100 000/año
Fracción atribuible poblacional	= 104,8/100 000/año ÷ 119,4/100 000/año = 0,88

[a]Datos de Thun MJ, Day-Lally CA, Calle EE, et al. Excess mortality among cigarette smokers: Changes in a 20-year interval. Am J Public Health 1995;85:1223-1230.

Riesgo atribuible

Puede que alguien se pregunte: «¿cuál es el riesgo adicional (incidencia) de la enfermedad tras la exposición en comparación con el riesgo que tienen las personas que no están expuestas?». La respuesta se expresa como **riesgo atribuible,** que es el riesgo absoluto (o incidencia) de la enfermedad en las personas expuestas menos el riesgo absoluto en las personas no expuestas. En la tabla 5-4 se muestra que el riesgo atribuible de muerte por cáncer de pulmón en los fumadores se calcula en 326,6 por 100 000 por año. El riesgo atribuible es la incidencia adicional de la enfermedad relacionada con la exposición, teniendo en cuenta la incidencia básica de la enfermedad debida a otras causas. Obsérvese que esta forma de comparar las tasas implica que el factor de riesgo es una causa, y no solamente un marcador. Por la forma en que se calcula, el riesgo atribuible también se conoce como **diferencia de riesgo,** es decir, la diferencia entre dos riesgos absolutos.

booksmedicos.org

Riesgo relativo

Por otro lado, puede que alguien se pregunte: «¿cuántas veces es más probable que las personas expuestas adquieran la enfermedad en comparación con las no expuestas?». Para responder a esta pregunta se emplea el **riesgo relativo** o **razón de riesgo,** es decir, el cociente entre la incidencia en las personas expuestas y la incidencia en las personas no expuestas estimado en la tabla 5-4 como 23,2. El riesgo relativo (o una estimación del riesgo relativo, la razón de momios *[odds ratio]*, que se expone en el capítulo 6) es el resultado que se notifica con más frecuencia en los estudios sobre riesgo, en parte por la comodidad de calcularlo, pero también porque constituye un sistema métrico común en los estudios de factores de riesgo similares, pero con tasas de incidencia basales diferentes. Puesto que el riesgo relativo indica la intensidad de la asociación entre exposición y enfermedad, se trata de una útil medida del efecto en los estudios etiológicos.

Interpretación del riesgo atribuible y el riesgo relativo

Aunque el riesgo atribuible y el riesgo relativo se calculan de los mismos dos componentes –la incidencia (o el riesgo absoluto) de un desenlace en un grupo expuesto y no expuesto–, el tamaño resultante del riesgo puede parecer muy diferente, dependiendo si se usa el riesgo atribuible o el relativo.

Ejemplo

Suponga que un factor de riesgo duplica la probabilidad de morir de una cierta enfermedad (p. ej., el riesgo relativo es 2). Luego, suponga que la frecuencia del factor de riesgo varía en cuatro grupos de personas no expuestas, de 1 en 10 en un grupo, hasta 1 en 10 000 en otro grupo, como se muestra a continuación. Los cálculos resultantes para el riesgo relativo y el riesgo atribuible serían:

booksmedicos.org

Incidencia		Riesgo relativo	Riesgo atribuible
Grupo no expuesto	Grupo expuesto		
1/10 000	2/10 000	2,0	0,2/1 000
1/1 000	2/1 000	2,0	2/1 000
1/100	2/100	2,0	20/1 000
1/10	2/10	2,0	200/1 000

Debido a que el cálculo del riesgo relativo anula la incidencia, no aclara el tamaño del riesgo en las conversaciones con los pacientes y no ayuda mucho en la decisión de tratar de modificar el riesgo en un paciente. Probablemente, el paciente y el médico estarán mucho más preocupados cuando la tasa de mortalidad se duplique de 100 a 200 en 1 000 que de 1 a 2 en 10 000. Un ejemplo de la vida real sobre el cambio en el riesgo absoluto, a pesar de que el riesgo relativo permanece estable, se muestra en un estudio sobre el riesgo de fractura en mujeres postmenopáusicas, considerando las mediciones basales de la densidad mineral ósea (DMO) y la edad (7). El riesgo relativo de una DMO baja cambió poco durante varias décadas (con un intervalo de 1,97-2,78), mientras que el riesgo atribuible casi se duplicó (14,4-23,8 por 1 000 mujeres-años), pues el riesgo de fracturas aumentó con la edad con independencia de la DMO (tabla 5-5).

En la mayor parte de las situaciones clínicas, lo más claro es simplemente concentrarse en el riesgo atribuible al comparar los riesgos absolutos en las personas expuestas y las no expuestas. Irónicamente, ya que la mayoría de la investigación clínica muestra sus resultados como un riesgo relativo, los médicos suelen centrarse en los riesgos relativos al aconsejar a los pacientes.

Riesgo de la población

Otra forma de concebir el riesgo es preguntarse: «¿en qué medida contribuye un factor de riesgo a las tasas globales de enfermedad en los grupos de personas en lugar de en pacientes concretos?». Esta información es útil para decidir qué factores de riesgo son especialmente importantes y qué factores son banales para la salud global de una comunidad, informando de cuáles son prioritarios para la implementación de los recursos sanitarios.

Para calcular el riesgo de la población, es necesario tener en cuenta la frecuencia con que los miembros de una comunidad se encuentran expuestos a un factor de riesgo. El **riesgo atribuible poblacional** es el producto del riesgo atribuible por la prevalencia de la exposición al factor

booksmedicos.org

de riesgo en una población, y mide la fracción de incidencia de la enfermedad en una comunidad que está asociada a un factor de riesgo. También se puede medir el exceso de incidencia de la enfermedad en una **población asociada a un factor de riesgo** concreto mediante la fracción atribuible en la población, que se obtiene dividiendo el riesgo atribuible en la población entre la incidencia total de la enfermedad en la población.

En el caso del tabaquismo y el cáncer de pulmón (tabla 5-4), el tabaquismo contribuye anualmente con unos 105 muertes por cáncer de pulmón por cada 100 000 hombres en la población (riesgo atribuible en la población) y representa el 88 % de todas las muertes por cáncer de pulmón en la población (fracción atribuible en la población). Considere la importancia de la prevalencia del factor de riesgo (tabaquismo) en estos cálculos. Conforme disminuyen las tasas de tabaquismo, la fracción de cáncer de pulmón causada por el tabaquismo también disminuye.

Tabla 5-5

Comparación de los riesgos relativos y atribuibles en la relación entre la puntuación T de la densidad mineral ósea, las fracturas y la edad

Edad	Incidencia de fracturas (riesgo absoluto por 1 000 personas-años)		Riesgo relativo[a]	Riesgo[b] atribuible por 1 000 personas-años
	Puntuación T DMO >–1,0	Puntuación T DMO <–2,0		
50-59	4,5	19,0	2,63	14,4
60-69	50,0	22,0	2,78	16,4
70-79	7,5	30,5	2,37	20,3
80-99	16,5	42,0	1,97	23,8

[a]Ajustado.
[b]Estimado de la figura 2 de referencia.
Datos de Siris ES, Brenneman SK, Barrett-Connor E, et al. The effect of age and bone mineral density on the absolute, excess, and relative risk of fracture in postmenopausal women aged 55-99: results from the National Osteoporosis Risk Assessment (NORA). Osteoporosis Int 2006;17:565-574.

booksmedicos.org

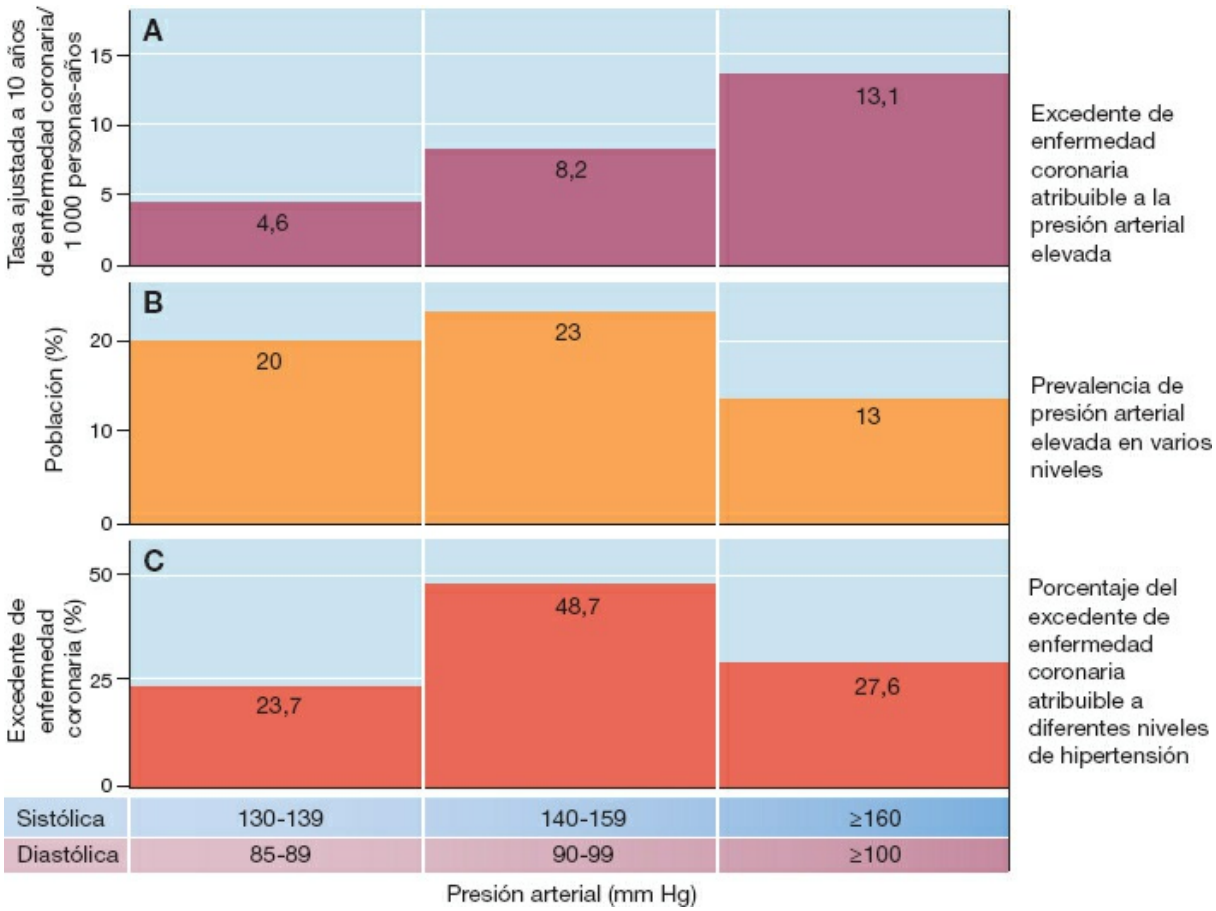


Figura 5-3 ■ Relaciones entre el riesgo atribuible, la prevalencia del factor de riesgo y la población de riesgo de muerte por enfermedad coronaria (EC) debido a la hipertensión. El cuadro A muestra que el riesgo atribuible para EC aumenta conforme se elevan los niveles de presión arterial. Sin embargo, debido a que la hipertensión leve y moderada son más prevalentes que la hipertensión severa (cuadro B), la mayoría del excedente de las muertes por EC causadas por hipertensión no se deben a los niveles más elevados de presión arterial (cuadro C). (Datos de Wilson PWF, Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97:1837-1847.)

Como se comentó en el capítulo 4, si un factor de riesgo relativamente débil es muy prevalente en una comunidad, podría ser responsable de más enfermedades que un factor de riesgo muy fuerte, pero infrecuente. La figura 5-3 ilustra este hecho para la hipertensión y el desarrollo de enfermedad coronaria. La figura 5-3 A muestra el riesgo atribuible (excedente) de enfermedad coronaria en base a diferentes niveles de hipertensión en un grupo de unos 2 500 hombres seguidos durante 10 años. El riesgo aumenta cuando se eleva la presión sanguínea. Sin embargo, pocos hombres tuvieron una presión arterial demasiado elevada (fig. 5-3 B). En consecuencia, el nivel más alto de hipertensión contribuyó sólo con

booksmedicos.org

un cuarto del excedente de la enfermedad coronaria en la población (fig. 5-3 C).

Por consiguiente, y de forma paradójica, los médicos podrían salvar más vidas con un tratamiento eficaz de los niveles más bajos de hipertensión arterial, en lugar de tratar los más altos. Este hecho, tan contrario a la creencia intuitiva del pensamiento clínico, se ha denominado «paradoja de la prevención» (8).

Las medidas de riesgo de la población se incluyen con menos frecuencia que las medidas de riesgo absoluto, atribuible y relativo, pero una consulta clínica concreta tiene el mismo valor de población para el médico que para los encargados de diseñar políticas sanitarias. Además, la forma en que la prevalencia de la exposición afecta al riesgo de la comunidad puede ser importante para la asistencia de pacientes concretos. Por ejemplo, cuando los pacientes no pueden proporcionar sus antecedentes o cuando les resulta difícil reconocer la exposición, los médicos dependen de la prevalencia de exposición habitual para calcular las probabilidades de varias enfermedades. Cuando se consideran las causas tratables de la cirrosis en un paciente estadounidense, es más útil pensar en el alcohol que en los esquistosomas, puesto que muy pocos estadounidenses están expuestos a *Schistosoma mansoni*. Obviamente, en el delta del Nilo, donde la esquistosomiasis es prevalente y los habitantes son en su mayoría musulmanes, por lo que es muy infrecuente que beban alcohol, se podría adoptar una postura muy diferente.

TENER EN CUENTA OTRAS VARIABLES

Hasta ahora hemos discutido principalmente sobre la exposición y la enfermedad de forma aislada, como si fueran las únicas dos variables importantes, aunque, de hecho, muchas otras variables son parte del fenómeno que se está estudiando; estas otras variables pueden tener uno de dos efectos importantes sobre los resultados. Pueden causar un cambio artificial e indeseable en la relación observada entre la exposición y la enfermedad (factor de confusión), causando conclusiones incorrectas sobre la relación o pueden modificar la magnitud de la relación entre la exposición y la enfermedad (modificación del efecto), que es una información valiosa para los médicos. En esta sección se discuten estos dos efectos, que son muy importantes para interpretar los resultados de una

booksmedicos.org

investigación, sobre todo en los estudios observacionales.

Variables extrañas

Variables extrañas es un término general para las variables que forman parte del sistema que se está estudiando pero que no son (es decir, son «ajenas» a) la exposición y la enfermedad de mayor interés. Por ejemplo, en un estudio sobre el ejercicio y la muerte súbita, las otras variables que son relevantes para el estudio son la edad, el IMC, la diabetes coexistente, todos los factores de riesgo cardiovascular y todo lo relacionado con la capacidad de ejercitar. Los estadísticos prefieren el término **covariable.** Ninguno de los términos es particularmente correcto. Las variables extrañas no son del todo «extrañas» porque pueden tener efectos importantes en la relación entre la exposición y la enfermedad. También las covariables pueden o no «covariar» (cambiar de acuerdo con la relación entre ellas, la exposición o la enfermedad), sin embargo, estos son los términos que se utilizan.

Descripciones simples sobre el riesgo

Los estudios observacionales pueden no tener en cuenta estas otras variables y comparar simplemente el curso de la enfermedad en dos grupos de origen natural, uno expuesto al factor de riesgo o pronóstico mientras que el otro no, sin considerar que la propia exposición es responsable de cualquier diferencia observada en el desenlace. Las **medidas crudas** del efecto (no ajustadas en base a otras variables) pueden ser útiles para predecir acontecimientos sin importar las causas.

Ejemplo

Investigadores británicos siguieron una cohorte de 17 981 niños diagnosticados con cáncer infantil antes de los 15 años de edad (9). Después de un seguimiento hasta los 66 años, hubo 11 veces más muertes que las que habrían ocurrido en la población general. El excedente en la tasa de mortalidad disminuyó con la edad, pero 45 años después del diagnóstico de cáncer, la tasa de mortalidad aún era 3 veces mayor que la tasa esperada. Las muertes se debieron principalmente a otros cánceres primarios y a enfermedades circulatorias. Los investigadores no reunieron evidencia a favor o en contra de las muchas razones posibles para el excedente en las tasas de mortalidad, como los efectos del tratamiento con radiación y la quimioterapia, el riesgo genético y los factores de riesgo ambiental. No obstante, es útil saber que las tasas de

booksmedicos.org

mortalidad son mucho más altas, así como las causas de muerte, para planificar la atención de los adultos que ha sobrevivido a un cáncer en la infancia.

En general, los investigadores quieren aportar más que las medidas crudas del efecto, y desean demostrar cómo se relaciona la exposición con la enfermedad, con *independencia* del resto de las variables que pueden afectar a esta relación. Es decir, quieren acercarse lo más posible a describir la causa y el efecto.

CONFUSIÓN

La validez de los estudios observacionales se ve amenazada sobre todo por la **confusión.** Ya hemos descrito la confusión en términos conceptuales en el capítulo 1, destacando que la confusión ocurre cuando la exposición se asocia o «viaja junto» a otra variable, que en sí misma se relaciona con el resultado, de modo que el efecto de la exposición puede confundirse con o distorsionarse por el efecto de la otra variable. La confusión causa un error sistemático –un sesgo– en la inferencia, por lo que los efectos de una variable se atribuyen a otra. Por ejemplo, en un estudio sobre si las vitaminas protegen contra episodios cardiovasculares, si las personas que eligen tomar vitaminas también son más propensas a seguir un estilo de vida saludable (p. ej., no fumar cigarrillos, ejercitar, comer una dieta adecuada y evitar la obesidad), tomar vitaminas se asociaría con tasas más bajas de ECV, independientemente de si las vitaminas protegen contra las ECV. La confusión puede aumentar o disminuir la asociación observada entre la exposición y la enfermedad.

Definición operativa

Una variable de confusión es aquella que:

- Se asocia con la exposición.
- Se asocia con la enfermedad.
- No es parte de la cadena causal de la exposición a la enfermedad.

La variable de confusión no puede estar en la cadena causal entre la exposición y la enfermedad; aunque las variables que están en la cadena necesariamente están relacionadas tanto con la exposición como con la

booksmedicos.org

enfermedad, no son los episodios iniciales. (En ocasiones, estas variables se denominan como **resultados intermedios**). Si sus efectos fueran eliminados, esto también eliminaría cualquier asociación que pudiera existir entre la exposición y la enfermedad. Por ejemplo, en un estudio sobre la dieta y la ECV, el colesterol sérico es consecuencia de la dieta; si el efecto del colesterol se eliminara, disminuiría de forma incorrecta la asociación entre la dieta y la ECV.

En la práctica, mientras que las **variables de confusión** (coloquialmente denominadas **factores de confusión**) pueden ser examinadas de una en una, por lo general muchas variables pueden confundir la relación entre la exposición y la enfermedad, por lo que todas son examinadas y controladas de forma concurrente.

Posibles factores de confusión

¿Cómo decidir qué variables deben ser consideradas *posibles* factores de confusión? Una estrategia es identificar todas las variables conocidas a partir de otros estudios que se asocian con la exposición o la enfermedad. La edad casi siempre es un candidato, pues es un factor de riesgo conocido para la enfermedad en cuestión. Otro método consiste en examinar las variables en los datos del estudio para determinar si existe asociación estadística con la exposición y la enfermedad, utilizando criterios liberales de «asociación» con el fin de no perder posibles factores de confusión. Los investigadores también pueden considerar variables que sólo tienen sentido de acuerdo con su experiencia clínica o con la biología de la enfermedad, sin importar si existen estudios de investigación sólidos que los vinculen con la exposición o la enfermedad. La intención es tender una red amplia con el fin de no perder posibles factores de confusión. Esto se debe a la posibilidad de que una variable pueda confundir la relación entre la exposición y la enfermedad por casualidad debido a los datos particulares disponibles, aunque no sea un factor de confusión natural.

Confirmación de la confusión

¿Cómo decidir si una variable que podría confundir la relación entre la exposición y la enfermedad realmente lo hace? Un método es demostrar que la variable se asocia con la exposición y (por separado) con la enfermedad. Otro método es determinar si la relación cruda entre la

booksmedicos.org

exposición y la enfermedad es diferente cuando se tiene en cuenta el posible factor de confusión. El siguiente ejemplo muestra ambos métodos.

Ejemplo

Como se señaló en un ejemplo del capítulo 4, varios estudios observacionales han demostrado que los niveles elevados de homocisteína en la sangre se relacionan con la ECV y que las vitaminas de folato disminuyen los niveles de homocisteína. Un estudio de cohorte examinó la relación entre la ingesta de folato y la incidencia de enfermedad cerebrovascular (10). El estudio inició con un total de 83 272 enfermeras de 34 a 59 años que se vigilaron durante 18 años con cuestionarios bienales para evaluar la dieta y la presencia de enfermedad cerebrovascular. El riesgo relativo de esta última, comparando el quintil más alto con el más bajo de la ingesta de folato, fue 0,83, lo que sugiere que el folato es un factor protector contra dicha patología. Sin embargo, la ingesta de folato se asoció inversamente con la ingesta de ácidos grasos trans y el tabaquismo (todos los factores de riesgo cardiovascular), y se asoció directamente con la probabilidad de hacer ejercicio (un factor protector) (fig. 5-4). Cuando se tienen en cuenta estos otros factores, el folato ya no fue un factor protector; el riesgo relativo fue 0,99. Los autores concluyeron que su estudio no mostró un efecto independiente del folato sobre la incidencia de enfermedad cerebrovascular; la relación cruda entre la exposición y la enfermedad ha sido confundida con otros factores de riesgo cardiovascular y factores protectores.

CONTROL DE LA CONFUSIÓN

Para determinar si un factor se relaciona independientemente al riesgo o al pronóstico, es ideal comparar las cohortes con y sin el factor, manteniendo todo lo demás igual. Pero en la vida real, «todo lo demás» generalmente no es igual en los estudios observacionales.

¿Qué se puede hacer con este problema? Existen varias formas posibles de **controlar**[†] las diferencias entre los grupos. Controlar es un término general para cualquier proceso destinado a retirar los efectos de variables extrañas mientras se examinan los efectos independientes de variables individuales. Se puede aplicar una variedad de métodos durante el diseño o el análisis de la investigación (se resumen en la tabla 5-6 y se describen en el siguiente texto). Deben aplicarse una o más de estas estrategias en cualquier estudio observacional que intente describir el efecto de una variable, independientemente de otras variables que puedan afectar al resultado. La pregunta básica es «¿Las diferencias entre los grupos de riesgo o pronóstico se relacionan con un factor particular del estudio o con

booksmedicos.org

algún otro factor/factores?».

[†]Desafortunadamente, el término «control» también tiene muchos otros significados: las personas no expuestas en un estudio de cohorte, los pacientes en un ensayo clínico que no reciben el tratamiento experimental y las personas sin la enfermedad (que no son casos) en un estudio de casos y controles.

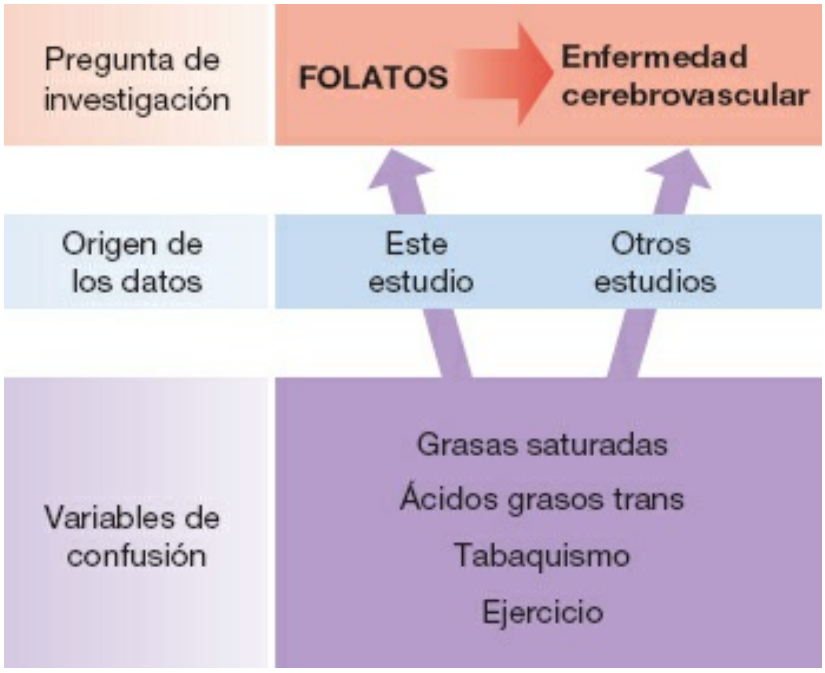


Figura 5-4 ■ Ejemplo de la confusión. La relación entre la ingesta de folatos y la incidencia de enfermedad cerebrovascular fue confundida por varios factores de riesgo cardiovascular y protectores.

Aleatorización

La mejor manera de equilibrar todas las variables extrañas entre los grupos es la asignación aleatoria de los pacientes en los grupos para que cada paciente tenga la misma probabilidad de caer en el grupo expuesto o no expuesto (v. cap. 9). Una característica especial de la aleatorización es que no sólo equilibra las variables que se sabe que afectan al resultado y las incluidas en el estudio, sino también los factores de confusión desconocidos o no medidos. Por desgracia, generalmente no es posible estudiar los factores de riesgo o pronóstico con ensayos aleatorizados.

Restricción

Los pacientes reclutados en un estudio pueden limitarse sólo a aquellos que

booksmedicos.org

poseen una estrecha gama de características, una estrategia denominada **restricción.** Cuando se adopta esta estrategia, ciertas características pueden ser similares en los grupos que van a compararse. Por ejemplo, el efecto de una ECV previa sobre el pronóstico tras un infarto de miocardio puede ser estudiado en pacientes sin antecedente de tabaquismo o hipertensión. Sin embargo, este método es limitado. Aunque la restricción sobre el ingreso al estudio sin duda puede producir grupos homogéneos de pacientes, lo hace a expensas de la generalización. Durante la exclusión de sujetos potenciales, las cohortes podrían ya no ser representativas de la mayoría de los pacientes con el trastorno. También, después de la restricción, ya no es posible aprender nadamás sobre los efectos de las variables excluidas en ese estudio.

Tabla 5-6

Métodos para controlar la confusión

Método	Descripción	Fase del estudio	
		Diseño	Análisis
Aleatorización	Asignar a los pacientes en grupos de manera que cada paciente tenga la misma probabilidad de estar en uno u otro grupo	+	
Restricción	Limitar el rango de las características de los pacientes en el estudio	+	
Pareamiento	Para cada paciente en un grupo, seleccione uno o más pacientes con las mismas características (excepto para la que se está estudiando) para un grupo de comparación	+	+
Estratificación	Comparar las tasas en subgrupos (estratos) que por lo demás tienen la misma probabilidad de presentar el resultado		+
Ajuste simple	Ajustar las tasas crudas matemáticamente para una o unas cuantas características para darle la misma ponderación a estratos de riesgo similar		+
Ajuste multivariado	Ajustar las diferencias en un gran número de factores relacionados con el resultado, usando técnicas de modelado matemático		+
Análisis del mejor caso/peor caso	Describe qué tan diferentes podrían ser los resultados bajo el supuesto más extremo (o simplemente muy poco probable) sobre el sesgo de selección		+

Tabla 5-7

Ejemplo de estratificación: tasas de muerte hipotéticas tras una cirugía de revascularización coronaria en dos hospitales, clasificados por el riesgo prequirúrgico

Riesgo preoperatorio	Hospital A			Hospital B		
	Pacientes	Muertes	Tasa (%)	Pacientes	Muertes	Tasa (%)
Alto	500	30	6	400	24	6
Medio	400	16	4	800	32	4
Bajo	300	2	0,67	1 200	8	0,67
Total	1 200	48	4	2400	64	2,7

booksmedicos.org

Pareamiento

El **pareamiento o emparejamiento** es otra forma de hacer que los pacientes de dos grupos sean similares. En su forma más simple, para cada paciente en el grupo expuesto, se seleccionarían uno o más pacientes con las mismas características (exceptuando el factor de interés) en el grupo de comparación. El pareamiento suele realizarse en las variables que están fuertemente relacionadas con el resultado, por lo que los investigadores quieren asegurarse que no son diferentes en los grupos que serán comparados. Con frecuencia, los pacientes se parean por edad y sexo, porque estas variables están fuertemente relacionadas con el riesgo o el pronóstico de muchas enfermedades, pero también puede ser útil el pareamiento para otras variables, como la etapa o la gravedad de la enfermedad y los tratamientos previos.

Aunque el pareamiento es frecuente y puede ser muy útil, tiene sus limitaciones. El pareamiento controla los sesgos sólo para las variables involucradas. Además, a menudo no es posible el pareamiento de más de unas pocas variables por las dificultades prácticas para encontrar pacientes que cumplan con todos los criterios del pareamiento. Por otro lado, si las categorías del pareamiento son relativamente crudas, puede haber lugar para diferencias sustanciales entre los grupos pareados. Por ejemplo, si las mujeres de un estudio sobre el riesgo de tener un hijo con síndrome de Down fueran pareadas según la edad materna con un intervalo de 10 años, podría haber una diferencia en la frecuencia de casi 10 veces relacionada con la edad, si la mayoría de las mujeres de un grupo tuvieran 30 años y la mayoría de las del otro grupo tuvieran 39 años. Por último, al igual que en la restricción, una vez que se parea de acuerdo con una variable, ya no pueden evaluarse sus efectos sobre los resultados en el estudio. Por estas razones, aunque se pueden parear algunas características que tienen una relación especialmente fuerte con el resultado, los investigadores también confían en otras formas para controlar el sesgo.

Estratificación

Con la **estratificación,** los datos se analizan y los resultados se presentan de acuerdo con subgrupos de pacientes, o estratos, con un riesgo o pronóstico similar (aparte de la exposición de interés). Un ejemplo de este método es el análisis de las diferencias en la mortalidad hospitalaria para

booksmedicos.org

un procedimiento quirúrgico común, la cirugía de derivación coronaria (tabla 5-7). Esto es especialmente relevante hoy en día debido a varios ejemplos importantes sobre las «calificaciones» para los médicos y los hospitales, y la preocupación que las diferencias reportadas puedan estar relacionadas con el paciente en vez de ser características del cirujano o del hospital.

Supongamos que queremos comparar las tasas de mortalidad operatoria de la cirugía de derivación coronaria en los hospitales A y B. En general, el hospital A informa de 48 muertes en 1 200 cirugías de derivación (4 %) y el hospital B refiere 64 muertes en 2 400 cirugías (2,6 %).

Las tasas crudas sugieren que el hospital B es superior. Pero, ¿realmente es superior si todo lo demás es igual? Quizás el riesgo preoperatorio entre los pacientes del hospital A fue mayor que en el hospital B y esto, en lugar de la atención hospitalaria, representó la diferencia entre las tasas de mortalidad. Para saber si esta posibilidad representa la diferencia observada en las tasas de mortalidad, los pacientes de cada uno de estos hospitales son agrupados en estratos con un riesgo preoperatorio similar en base a la edad, la función previa del miocardio, la extensión de la enfermedad oclusiva y otras características. Después se comparan las tasas de mortalidad operatoria en cada estrato de riesgo.

La tabla 5-7 muestra que cuando se divide a los pacientes de acuerdo con el riesgo preoperatorio, las tasas de mortalidad operatoria en cada estrato de riesgo son idénticas en los dos hospitales: 6 % en los pacientes de riesgo alto, 4 % en los pacientes con riesgo moderado y 0,67 % en los pacientes con riesgo bajo. Las tasas crudas eran engañosas debido a diferencias importantes en las características de riesgo en los pacientes tratados en los dos hospitales: 42 % de los pacientes del hospital A y sólo un 17 % de los pacientes del hospital B tuvieron alto riesgo.

Una ventaja de la estratificación es que se trata de una forma relativamente transparente de reconocer y controlar los sesgos.

Estandarización

Si una variable extraña tiene una relación especialmente fuerte con los resultados, es posible comparar dos tasas sin sesgos relacionados con esta variable si se ajustan, para igualar el peso que se le ha dado a esa variable. Este proceso, denominado **estandarización** (o ajuste), muestra cuál sería la

booksmedicos.org

tasa global para cada grupo si se aplicaran tasas específicas para cada estrato en una población compuesta por proporciones similares de personas en cada estrato.

Para ejemplificar este proceso, suponga que la mortalidad operatoria en los hospitales A y B puede ajustarse a una distribución común de grupos de riesgo, asignando a cada estrato de riesgo la misma ponderación en los dos hospitales. Sin el ajuste, los estratos de riesgo recibirían diferente ponderación en los dos hospitales. La tasa de mortalidad del 6 % para pacientes de alto riesgo recibe una ponderación de 500/1 200 en el hospital A y una ponderación mucho menor de 400/2 400 en el hospital B. El resultado es una tasa cruda para el hospital A, que es la suma de la tasa en cada estrato multiplicado por la ponderación: (500/1 200 × 0,06) + (400/1 200 × 0,04) + (300/1 200 × 0,0067) = 0,04. De forma similar, la tasa cruda para el hospital B es (400/2 400 × 0,06) + (800/2 400 × 0,04) + (1 200/2 400 × 0,0067) = 0,027.

Si utilizáramos la misma ponderación para comparar los dos hospitales, la comparación sería justa (libre del efecto de las diferentes proporciones en los diferentes grupos de riesgo). La decisión de ponderar no es significativa siempre y cuando sea la misma en los dos hospitales. La ponderación podría basarse en la que está presente en cualquiera de los dos hospitales o en una población de referencia. Por ejemplo, si a cada estrato se le asignara una ponderación de 1/3, la tasa estandarizada para el hospital A = (0,33 × 0,06) + (0,33 × 0,04) + (0,33 × 0,0067) = 0,035, que es exactamente igual a la tasa estandarizada para el hospital B. La consecuencia de asignar la misma ponderación a los estratos de cada hospital es eliminar el riesgo aparentemente excesivo del hospital A.

La estandarización se suele utilizar en comparaciones relativamente crudas para ajustar en base a una sola variable como la edad, que obviamente difiere entre los grupos que serán comparados. Por ejemplo, los resultados crudos en el ejemplo del folato/enfermedad cerebrovascular se ajustaron según la edad, como se comentó antes. La estandarización tiene menor utilidad como una estrategia independiente para controlar la confusión cuando se deben considerar múltiples variables.

Ajuste multivariado

En la mayoría de las situaciones clínicas, muchas variables actúan entre sí para producir efectos, manteniendo relaciones complejas. Pueden estar

booksmedicos.org

relacionadas entre sí y con el resultado de interés. El efecto de una podría modificarse por la presencia de otras y los efectos conjuntos de dos o más variables pueden ser mayores que sus efectos individuales agrupados.

El **análisis multivariado** permite considerar los efectos de muchas variables de manera simultánea. Otros términos para este método incluyen el modelado matemático y el ajuste multivariable. El modelado se utiliza para ajustar (controlar) los efectos de muchas variables de manera simultánea para determinar los efectos independientes de cada una. Este método también puede seleccionar, de una larga lista de variables, aquellas que contribuyen de manera independiente a la variación global del resultado. El modelado también puede ordenar las variables de acuerdo con la fuerza de su contribución. Existen numerosos tipos de modelos prototípicos de acuerdo con el diseño y los datos en el estudio. Los estudios de cohortes y casos y controles en general dependen de la **regresión logística,** que se usa específicamente para resultados dicotómicos. El **modelo de riesgo proporcional de Cox** se utiliza cuando el resultado es el tiempo esperado para que ocurra un acontecimiento, como en los análisis de supervivencia (v. cap. 7).

En los estudios observacionales, el análisis multivariable es el único método factible para controlar muchas variables simultáneas durante la fase de análisis del estudio. La aleatorización también controla muchas variables, pero durante las fases de diseño y desarrollo del estudio. El pareamiento puede controlar sólo unas pocas variables a la vez y el análisis estratificado de múltiples variables corre el riesgo de tener muy pocos pacientes en algunos estratos. La desventaja del modelado es que para la mayoría de nosotros es una «caja negra», por lo que es difícil reconocer cuándo podría ser engañoso este método. Cuando se utiliza de forma adecuada, el modelado se utiliza en conjunto con, no en lugar de, el pareamiento y el análisis estratificado.

Estrategia general para controlar la confusión

Con excepción de la aleatorización, todas las formas para tratar con las diferencias entre las variables extrañas de los grupos comparten una limitación: son eficaces sólo para las variables seleccionadas que fueron consideradas. No tratan con los factores de riesgo o pronóstico que no se conocen en el momento del estudio ni con las que se conocen pero que no se tomaron en consideración.

booksmedicos.org

Por esta razón, además de las fortalezas y debilidades complementarias de varios métodos, no debemos depender solamente de uno u otro método para controlar los sesgos, sino usar varios métodos en conjunto, empalmados uno en el otro.

Ejemplo

En un estudio sobre la asociación de las contracciones ventriculares prematuras con una menor supervivencia en los años siguientes a un infarto agudo de miocardio, podríamos hacer frente a la confusión de la siguiente manera:

- *Restringir* el estudio a los pacientes que no son muy viejos o jóvenes y cuyo infarto no se debió a una causa inusual, como arteritis o aneurisma disecante.
- *Parear* de acuerdo con la edad, un factor fuertemente relacionado con el pronóstico pero extraño a la pregunta principal.
- Examinar los resultados separándolos por estratos con diferente gravedad clínica utilizando el *análisis estratificado*. Esto incluye la presencia o ausencia de insuficiencia cardíaca congestiva u otras enfermedades, como la enfermedad pulmonar obstructiva crónica.
- Utilizando el *análisis multivariado*, ajustar los resultados crudos según los efectos de todas las variables además de la arritmia, que en conjunto podrían estar relacionadas con el pronóstico.

ESTUDIOS OBSERVACIONALES Y CAUSA

El resultado final de un estudio observacional cuidadoso, controlado para una gran cantidad de variables extrañas, es acercarse tanto como sea posible a describir un efecto independiente verdadero, uno separado del resto de las variables que confunden la exposición, la relación con la enfermedad. Sin embargo, siempre es posible que otras variables importantes no se hayan tomado en cuenta, ya sea porque no se conocía su importancia o porque no fueron o no podían ser medidas. La consecuencia de los factores de **confusión no medidos** es la **confusión residual.** Por esta razón se debe considerar que un estudio único (y los investigadores deben describir sus resultados) establece «asociaciones independientes» y no necesariamente una causa. En el capítulo 12 se describe cómo desarrollar un estudio para encontrar la asociación causal.

booksmedicos.org

MODIFICACIÓN DEL EFECTO

Una situación muy diferente de la confusión es si la presencia o ausencia de una variable cambia el efecto de la exposición sobre la enfermedad, lo que se conoce como **modificación del efecto.** Como afirma Rothman (11),

> La diferencia más importante es que, mientras la confusión es un sesgo que el investigador espera prevenir o, de ser necesario, eliminar de los datos, la modificación del efecto es una descripción detallada del propio efecto. Por lo tanto, la modificación del efecto es un hallazgo que debe ser reportado en lugar de un sesgo que debe ser evitado. El análisis epidemiológico generalmente está destinado a eliminar la confusión y a descubrir y describir la modificación del efecto.

Los estadísticos llaman **interacción** a la modificación del efecto, mientras que los biólogos la conocen como sinergia o antagonismo, dependiendo si este tercer factor aumenta o disminuye el efecto.

booksmedicos.org

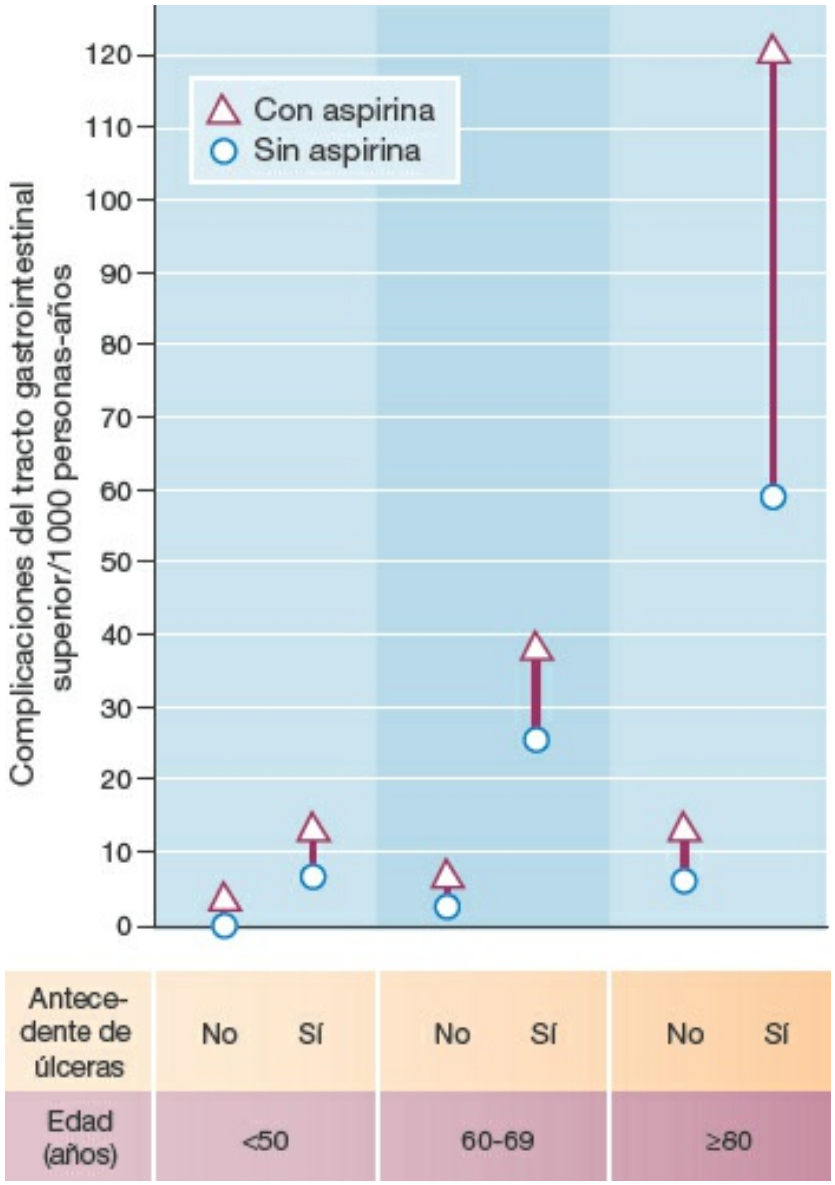


Figura 5-5 ■ **Ejemplo de modificación del efecto.** El riesgo adicional de complicaciones gastrointestinales causadas por la aspirina está modificado por la edad y los antecedentes de enfermedad ulcerosa péptica. (Datos de Patrono C, Rodriguez LAG, Landolfi R, et al. Low-dose aspirin for the prevention of atherothrombosis. N Engl J Med 2005;353: 2373-2383.)

Ejemplo

Se ha demostrado que la aspirina previene los episodios cardiovasculares. Su recomendación depende del riesgo de estos episodios en un paciente y de las complicaciones de la aspirina, principalmente de la hemorragia gastrointestinal alta. La figura 5-5 muestra que la tasa adicional de complicaciones gastrointestinales (incidencia atribuible a la aspirina) depende de otros dos factores, la edad y el antecedente de enfermedad ulcerosa péptica (12). En hombres

booksmedicos.org

menores de 50 años sin antecedentes de úlcera, la tasa de complicaciones es de alrededor de 1/1 000 persona-años y virtualmente no hay riesgo adicional relacionado con la aspirina. El riesgo aumenta un poco con la edad entre los hombres sin antecedentes de úlcera. Sin embargo, entre los que tienen el antecedente de enfermedad ulcerosa, la tasa de complicaciones aumenta demasiado con la edad, y aún más con el uso de aspirina. En el nivel más alto de riesgo, en hombres mayores de 80 años con antecedente de úlcera, la aspirina duplica el riesgo de 60 a 120/1 000 persona-años.

Este ejemplo muestra que la edad y el antecedente de enfermedad ulcerosa péptica modifican el efecto de la aspirina sobre las complicaciones gastrointestinales. La información adicional proporcionada por la modificación del efecto permite que los médicos ajusten sus recomendaciones sobre el uso de aspirina acercándose más a las características de cada paciente.

La confusión y la modificación del efecto son independientes el uno del otro. Una variable determinada puede ser un factor de confusión, modificador del efecto, ambos o ninguno, dependiendo de la pregunta de investigación y los datos.

Preguntas de revisión

Seleccione la mejor respuesta para la pregunta 5-1.

5-1. ¿Cuál de las siguientes afirmaciones no es correcta tanto para los estudios de cohorte prospectivos como los retrospectivos?

A. Miden directamente la incidencia de la enfermedad.

B. Permiten evaluar las posibles asociaciones entre la exposición y muchas enfermedades.

C. Permiten que los investigadores decidan de antemano los datos que van a registrar.

D. Evitan los sesgos que pueden ocurrir si la exposición se mide después de conocer el resultado de interés.

Las preguntas 5-2 a 5-4 se basan en el siguiente ejemplo:

Se realizó un estudio para examinar la relación entre el tabaquismo, la

booksmedicos.org

enfermedad cerebrovascular y la edad (13). En un período de 12 años, la incidencia de enfermedad cerebrovascular en 1 000 personas (riesgo absoluto) de acuerdo con la edad y el hábito tabáquico fue:

Edad	No fumadores	Fumadores
45-49	7,4	29,7
65-69	80,2	110,4

5-2. ¿Cuál fue el riesgo relativo de enfermedad cerebrovascular entre los fumadores en comparación con los no fumadores a sus 40 años?

A. 1,4.
B. 4,0.
C. 22,3.
D. 30,2.
E. 72,8.
F. 80,7.

5-3. ¿Cuál fue el riesgo atribuible de enfermedad cerebrovascular por cada 1 000 personas entre los fumadores en comparación con los no fumadores a sus 60 años?

A. 1,4.
B. 4,0.
C. 22,3.
D. 30,2.
E. 72,8.
F. 80,7.

5-4. ¿Cuál de las siguientes afirmaciones sobre el estudio es incorrecta?

A. Se requiere información adicional para calcular el riesgo atribuible a la población de fumar entre las personas de 60 años.
B. Se presentaron más casos de enfermedad cerebrovascular debido al tabaquismo en personas de 60 años que en personas de 40 años.
C. Cuando se calcula el riesgo relativo, los resultados reflejan la información sobre la incidencia en las personas expuestas y no

booksmedicos.org

expuestas, mientras que los resultados del riesgo atribuible no lo hacen.

D. El riesgo relativo calculado es un argumento más fuerte para el tabaquismo como causa de enfermedad cerebrovascular para las personas en sus 40 años que el riesgo calculado para las personas en sus 60 años.

E. Dependiendo de la pregunta formulada, en el estudio se puede considerar que la edad es una variable que confunde o modifica el efecto.

Las preguntas 5-5 a 5-11 se basan en el siguiente ejemplo:

La trombosis venosa profunda (TVP) es un trastorno serio que ocasionalmente produce embolia pulmonar y la muerte (14). La incidencia de TVP aumenta debido a varios factores genéticos y ambientales, incluyendo los anticonceptivos orales (ACO) y una mutación genética, el factor V de Leiden. Estos dos factores de riesgo, los ACO y el factor V de Leiden interactúan. Los heterocigotos para el factor V de Leiden tienen un riesgo de TVP 4 a 10 veces mayor que la población general. En las mujeres sin la mutación genética, la incidencia de TVP incrementa un 0,8/10 000 mujeres/año en las que no toman ACO a 3,0/10 000 mujeres/año que toman ACO. La incidencia basal de TVP en los heterocigotos para el factor V de Leiden es 5,7/10 000 mujeres/año y aumenta a 28,5/10 000 mujeres/año en las que toman ACO. Las mutaciones del factor V de Leiden ocurren en un 5 % de los caucásicos, pero está ausente en africanos y asiáticos.

Seleccione la mejor respuesta para las preguntas 5-5 a 5-10.

5-5. ¿Cuál es el riesgo absoluto de TVP en las mujeres que no tienen la mutación y no toman ACO?

A. 0,8/10 000/año.

B. 1,3/10 000/año.

C. 2,2/10 000/año.

D. 9,5/10 000/año.

E. 25,5/10 000/año.

booksmedicos.org

5-6. ¿Cuál es el riesgo atribuible de TVP para las mujeres que toman ACO y no son portadoras de la mutación del factor V de Leiden en comparación con las que no toman ACO y no son portadoras de la mutación?

A. 0,8/10 000/año.
B. 1,3/10 000/año.
C. 2,2/10 000/año.
D. 9,5/10 000/año.
E. 2,5/10 000/año.

5-7. ¿Cuál es el riesgo atribuible de TVP en las mujeres que toman ACO y son portadoras del factor V de Leiden en comparación con las que toman ACO, pero que no son portadoras de la mutación?

A. 0,8/10 000/año.
B. 1,3/10 000/año.
C. 2,2/10 000/año.
D. 9,5/10 000/año.
E. 25,5/10 000/año.

5-8. En una población de 100 000 mujeres caucásicas usuarias de ACO, ¿cuál es el riesgo atribuible a la población para TVP en las mujeres heterocigotas para el factor V de Leiden?

A. 0,8/10 000/año.
B. 1,3/10 000/año.
C. 2,2/10 000/año.
D. 9,5/10 000/año.
E. 25,5/10 000/año.

5-9. ¿Cuál es el riesgo relativo de TVP en mujeres que toman ACO y son heterocigotas para el factor V de Leiden en comparación con las que toman ACO, pero que no son portadoras de la mutación?

A. 3,8.
B. 7,1.
C. 9,5.

booksmedicos.org

D. 28,5.
E. 35,6.

5-10. ¿Cuál es el riesgo relativo de TVP en mujeres que toman ACO y no tienen la mutación en comparación con las mujeres sin la mutación que no toman ACO?

A. 3,8.
B. 7,1.
C. 9,5.
D. 28,5.
E. 35,6.

5-11. Considerando la información de este estudio y los cálculos para las preguntas 5-5 a 5-10, ¿cuál de las siguientes afirmaciones sobre el riesgo de desarrollar TVP es incorrecta?

A. El factor V de Leiden modifica el efecto de los ACO sobre el riesgo anual de desarrollar TVP al aumentar el riesgo de 3 por cada 10 000 a cerca de 30 por cada 10 000 mujeres.
B. Los heterocigotos para el factor V de Leiden casi duplican el riesgo de TVP si se ingieren ACO.
C. A las mujeres heterocigotas para el factor V de Leiden se les debe aconsejar que no tomen ACO debido al alto riesgo relativo para TVP entre estas mujeres.

Seleccione la mejor respuesta para la pregunta 5-12.

5-12. En un estudio para determinar si tomar regularmente aspirina previene la muerte cardiovascular, las personas que la tomaron murieron con la misma frecuencia que los que no la tomaron. Sin embargo, los que la tomaron estaban más enfermos y tenían enfermedades más propensas a ser tratadas con aspirina. ¿Cuál de los siguientes métodos es la mejor manera de explicar la propensión de las personas a tomar aspirina?

A. Calcular el riesgo absoluto de muerte cardiovascular en los dos grupos y la diferencia del riesgo atribuible al uso de aspirina.
B. Crear subgrupos de pacientes que toman o no aspirina con

booksmedicos.org

indicaciones similares para usar el medicamento y comparar las tasas de muerte entre los subgrupos.

C. Parear a cada persona que toma aspirina con una que no la toma según la edad, el sexo y la comorbilidad, y comparar las tasas muerte en dos grupos.

BIBLIOGRAFÍA

1. Dawber TR. The Framingham Study: The Epidemiology of Atherosclerotic Disease. Cambridge, MA: Harvard University Press; 1980.
2. Kannel WB, Feinleib M, McNamara PM, et al. An investigation of coronary heart disease in families. The Framingham Offspring Study. *Am J Epidemiol* 1979;110:281–290.
3. Wen CP, Wai J PM, Tsai MK, et al. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. *Lancet* 2011;378:1244–1253.
4. Madsen KM, Hviid A, Vestergaard M, et al. A population-based study of measles, mumps, and rubella vaccination and autism. *N Engl J Med* 2002;347:1477–1482.
5. The Editors of The Lancet. Retraction—ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* 2010;375:445.
6. Geiger AM, Yu O, Herrinton LJ, et al. (on behalf of the CRN PROTECTS Group). A case-cohort study of bilateral prophylactic mastectomy efficacy in community practices. *Am J Epidemiol* 2004;159:S99.
7. Siris ES, Brenneman SK, Barrett-Connor E, et al. The effect of age and bone mineral density on the absolute, excess, and relative risk of fracture in postmenopausal women aged 55–99: results from the National Osteoporosis Risk Assessment (NORA), *Osteoporosis Int* 2006; 17:565–574.
8. Hofman A, Vandenbroucke JP. Geoffrey Rose's big idea. *Br Med J* 1992;305:1519–1520.
9. Reulen RC, Winter DL, Frobisher C, et al. long-term cause-specific mortality among survivors of childhood cancers. *JAMA* 2010;304:172–179.
10. Al-Delaimy WK, Rexrode KM, Hu FB, et al. Folate intake and risk of stroke among women. *Stroke* 2004;35: 1259–1263.
11. Rothman KJ. Modern Epidemiology. Boston: Little Brown and Co.; 1986.
12. Patrono C, Rodriguez LAG, Landolfi R, et al. Low-dose aspirin for the prevention of atherothrombosis. *N Engl J Med* 2005;353:2373–2383.
13. Psaty BM, Koepsell TD, Manolio TA, et al. Risk ratios and risk differences in estimating the effect of risk factors for cardiovascular disease in the elderly. *J Clin Epidemiol* 1990;43: 961–970.
14. Vandenbroucke JP, Rosing J, Bloemenkemp KW, et al. Oral contraceptives and the risk of venous thrombosis. *N Engl J Med* 2001;344:1527–1535.

booksmedicos.org

«... tome dos grupos que considere que representan a las personas que tienen o no la enfermedad y determine el porcentaje de cada grupo que presenta la característica… Ello produce, no una tasa verdadera, sino en su lugar, lo que generalmente se conoce como frecuencia relativa»

—Jerome Cornfield
1952

PALABRAS CLAVE

Período de latencia
Estudio de casos y controles
Control
Estudio de casos y controles basado en la población
Estudio de casos y controles anidado
Pareamiento
Pareamiento en sombrilla
Sobrepareamiento
Sesgo de recuerdo
Razón de momios
Riesgo relativo estimado
Razón de momios de prevalencia
Razón de momios cruda
Razón de momios ajustada
Curva epidémica

Los estudios de cohorte son una forma directa y maravillosamente lógica de

booksmedicos.org

estudiar el riesgo, si bien presentan limitaciones de carácter práctico. La mayor parte de las enfermedades crónicas tardan mucho en aparecer. El **período de latencia,** o período que transcurre entre la exposición a un factor de riesgo y la expresión de sus efectos patológicos, se mide en decenios en la mayor parte de las enfermedades crónicas. Por ejemplo, el tabaquismo precede a la enfermedad coronaria, al cáncer de pulmón y a la bronquitis crónica en 20 años o más, y la osteoporosis con fracturas que sufren los ancianos es consecuencia de los patrones alimentarios y de ejercicio físico adoptados a lo largo de la vida. Asimismo, en una cohorte, un número relativamente reducido de personas presentaba el resultado de interés, aunque es necesario que la medición de la exposición y el seguimiento se realicen en todos los miembros de la misma. El resultado es que los estudios de cohorte sobre riesgo precisan de mucho tiempo y esfuerzo, por no mencionar el dinero, para dar con una respuesta. La ineficacia es especialmente limitante en las enfermedades muy infrecuentes.

Algunas de estas limitaciones pueden superarse modificando los métodos de estudio de las cohortes, como en el caso de los diseños de casos y cohortes descritos en el capítulo anterior. Este capítulo describe otro modo de estudiar la relación entre un posible factor de riesgo (o de protección) y una enfermedad de forma más eficiente: mediante los estudios de casos y controles. Este método presenta dos ventajas principales sobre los estudios de cohorte: una, supera la necesidad de recoger datos de un gran número de personas, la mayoría de las cuales no contraen la enfermedad y, por lo tanto, contribuirá poco con los resultados, y dos, se trata de un método más rápido, ya que no es necesario esperar a que, tras la medición de la exposición, se produzca un efecto.

Sin embargo, la eficacia y el ahorro de tiempo tienen un costo: en los estudios de casos y controles, el manejo del sesgo se convierte en una tarea más difícil y a veces incierta. Además, estos estudios sólo producen una estimación del riesgo relativo y ninguna información directa sobre las restantes medidas del efecto, como el riesgo absoluto, el riesgo atribuible y los riesgos de la población que se describieron en el capítulo anterior. En la tabla 6-1 se resumen las ventajas e inconvenientes de los estudios de cohorte, de casos y controles.

Lo más importante es decidir entre la fortaleza científica y la factibilidad. De hecho, estos estudios son indispensables para el estudio del riesgo de enfermedades muy infrecuentes.

booksmedicos.org

Tabla 6-1

Resumen de las características de los estudios de cohorte y de casos y controles

Estudio de cohorte	Estudio de casos y controles
Comienza con una cohorte bien definida	Comienza con una muestra de casos y controles
La exposición se mide en los miembros de la cohorte	La exposición se mide en los casos y controles, algunas veces después de los resultados
Los casos surgen en la cohorte durante el seguimiento	La exposición ocurre antes de que las muestras se conviertan en casos y controles
La incidencia se mide en los miembros expuestos y no expuestos de la cohorte	La exposición se mide en los casos y controles
Se pueden calcular directamente los riesgos absoluto, relativo, atribuible y poblacional	Se puede estimar el riesgo relativo, pero no existe información sobre la incidencia

Ejemplo

A mediados de la década de 2000, los médicos comenzaron a reportar casos de una forma infrecuente de fractura femoral en las mujeres. Se pensó que los bisfosfonatos, unos fármacos prescritos para prevenir la osteoporosis, eran la causa, pues se habían introducido en las décadas anteriores y actuaban reduciendo la remodelación ósea. Las series de casos proporcionaban una asociación entre los bisfosfonatos y las fracturas atípicas, pero las mujeres en estos estudios tomaban otros fármacos y tenían otras enfermedades que también podrían haber estado relacionadas con el riesgo de fracturas. Para obtener una respuesta más definitiva sobre si los bisfosfonatos se asociaban de forma independiente con las fracturas atípicas, investigadores suecos realizaron un estudio de casos y controles (1). A partir del registro nacional de pacientes en Suecia identificaron a todas las 59 mujeres de 55 años o mayores que habían tenido fracturas femorales atípicas en 2008. También identificaron a 263 controles, mujeres en el mismo registro que tuvieron fracturas femorales ordinarias (para parearlas de acuerdo con la vulnerabilidad subyacente a las fracturas). Se registraron otras variables que podrían haber estado relacionadas con el uso de bisfosfonatos y fracturas atípicas, como la edad, el uso de fármacos modificadores óseos como los corticoesteroides o los estrógenos y las enfermedades como osteoporosis y fracturas previas. Después de tener en cuenta todos estos otros factores, las mujeres que tomaron

booksmedicos.org

bisfosfonatos tuvieron una probabilidad 33 veces mayor de presentar fracturas atípicas.

Al añadir un grupo de comparación y considerar otras variables que podrían haber estado relacionadas con el uso de bisfosfonatos y fracturas atípicas, los investigadores fueron capaces de deducir que los bisfosfonatos podrían ser causa de fracturas atípicas mucho más allá de lo que era posible sólo con una serie de casos.

Este capítulo, el tercero sobre riesgo, se titula «Riesgo: de la enfermedad a la exposición» porque los estudios de casos y controles implican retroceder desde la enfermedad a la exposición, a diferencia de los estudios de cohorte, que se proyectan hacia delante, desde la exposición a la enfermedad.

ESTUDIOS DE CASOS Y CONTROLES

En la figura 6-1 se presenta el diseño básico de un **estudio de casos y controles** en forma de diagrama. Se seleccionan dos muestras: una de ellas con pacientes que han desarrollado la enfermedad en cuestión, y otra formada por personas que no la han presentado. A continuación los investigadores retroceden en el tiempo para medir la frecuencia de exposición a un posible factor de riesgo en ambos grupos. Los datos resultantes pueden utilizarse para calcular el riesgo relativo de una enfermedad relacionado con el posible factor de riesgo.

Ejemplo

Los traumatismos craneoencefálicos son relativamente frecuentes entre los esquiadores alpinos y los que practican snowboard. Parece posible que los cascos pueden prevenir estos traumatismos, pero los detractores señalan que los cascos también podrían incrementar los traumatismos craneoencefálicos al reducir el campo de visión, alterar la audición y ofrecer a los atletas una falsa sensación de seguridad. Para obtener evidencia más definitiva sobre los efectos reales de los cascos, investigadores noruegos realizaron un estudio de casos y controles (fig. 6-2) (2). Se eligieron los casos y controles entre los visitantes de ocho centros principales de esquí alpino en hoteles de Noruega durante la temporada de invierno de 2002. Los casos fueron todas las 578 personas con traumatismos craneoencefálicos proporcionadas por la patrulla alpina. Los controles fueron una muestra de personas que esperaban en una fila en la base del teleférico principal de esquí en cada uno de los 8 hoteles. Tanto en los casos como en los controles, los investigadores registraron otros factores que podrían haber confundido la relación entre el uso de casco y el traumatismo craneoencefálico, incluyendo la edad, el sexo, la nacionalidad, el tipo de equipo, la asistencia previa a una escuela de esquí, si el equipo era propio o rentado y la habilidad para

booksmedicos.org

esquiar. Después de considerar los factores de confusión, el uso de casco se asoció con una reducción del 60 % sobre el riesgo de traumatismo craneoencefálico.

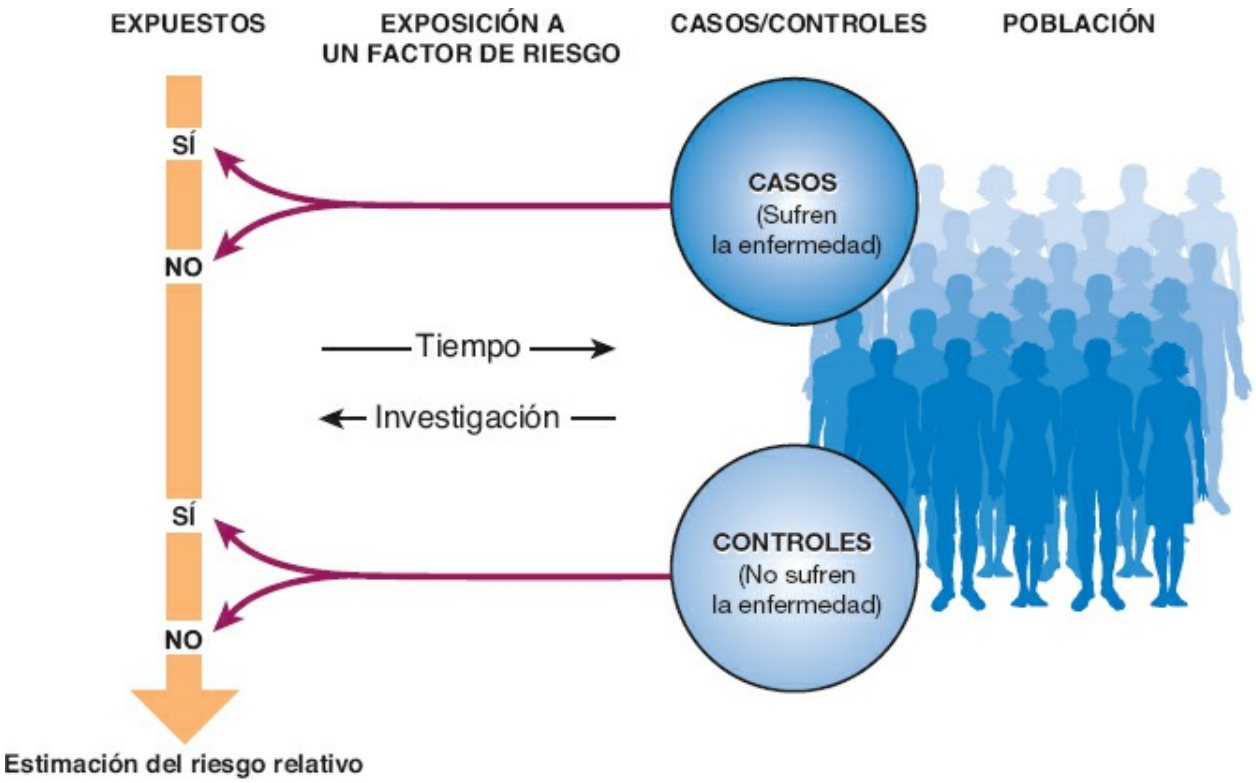


Figura 6-1 ■ Diseño de los estudios de casos y controles.

booksmedicos.org

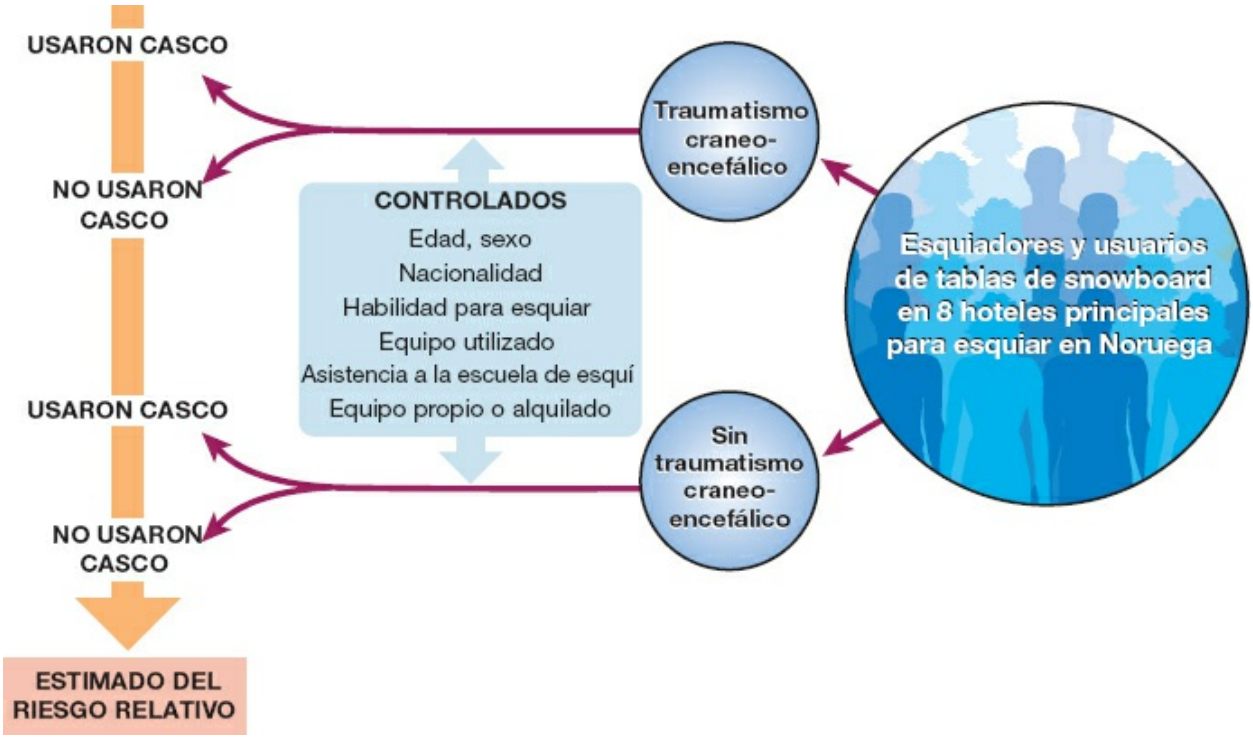


Figura 6-2 ■ Un estudio de casos y controles sobre el uso de casco y traumatismos craneoencefálicos entre los esquiadores y los usuarios de tabla de snowboard. (Resumen de Sulheim S, Holme I, Ekeland A, et al. Helmet use and risk of head injuries in alpine skiers and snowboarders. JAMA 2006;295:919-924.)

El término **control** también se emplea en otras situaciones. Se utiliza en estudios experimentales para referirse a personas, animales o material biológico que no han recibido la intervención del estudio. En los laboratorios de análisis clínicos, los «controles» son las muestras que contienen una cantidad conocida del material en estudio. Como verbo, *controlar* se utiliza para describir el proceso de tener en cuenta, neutralizar o restar los efectos de las variables extrañas a la pregunta principal. En este capítulo, el término se utiliza en el contexto de los estudios de casos y controles para referirse a las personas que no tienen la enfermedad o el resultado que se está estudiando.

DISEÑO DE LOS ESTUDIOS DE CASOS Y CONTROLES

La validez de los estudios de casos y controles depende de la minuciosidad con que se seleccionen los casos y los controles, de cómo se mida la exposición y de cómo se controlen las variables extrañas.

Selección de casos

booksmedicos.org

Los casos en un estudio de casos y controles deberían ser nuevos (incidentes), no los existentes (prevalentes). Las razones para ello se basan en los conceptos tratados en el capítulo 2. La prevalencia de una enfermedad en un momento del tiempo está en función de la incidencia y la duración de dicha enfermedad. La duración, a su vez, viene determinada por la proporción de pacientes que abandonan el estado de enfermedad (porque se recuperan o fallecen) o persisten en dicho estado (a causa de una evolución lenta o de un proceso paliativo satisfactorio). Como consecuencia de estas relaciones, los factores de riesgo de una enfermedad prevalente pueden constituir factores de riesgo de la incidencia, de la duración o de ambas; las aportaciones relativas de las dos medidas no pueden determinarse. Como ejemplo, si se estudiaran los casos prevalentes, una exposición que provocara una forma de enfermedad con una rápida mortalidad haría descender los casos que estuvieran expuestos, lo que reduciría el riesgo relativo y, por tanto, sugeriría que la exposición es menos nociva de lo que es realmente, o incluso que se trata de una exposición protectora.

Lo ideal sería que un estudio de casos y controles incluyera todos los casos o una muestra representativa de los que surjan en una población definida. Por ejemplo, en el estudio de los bisfosfonatos se incluyeron a todos los residentes de Suecia en 2008 y, en el estudio sobre los cascos, se contempló a todos los esquiadores y practicantes de snowboard en 8 hoteles principales de Noruega (que representan el 55 % de las pistas de esquí en el país).

Algunos estudios de casos y controles, especialmente los más antiguos, han identificado casos en hospitales y centros de referencia, donde es más probable encontrar enfermedades poco comunes. Esta forma de elegir los casos resulta cómoda, pero entraña problemas de validez. Estos centros pueden atraer casos particularmente graves o atípicos o aquellos con exposiciones inusuales—la muestra será incorrecta si la pregunta de investigación subyacente es sobre enfermedades y exposiciones ordinarias en los estudios de casos y controles.

También, en esta situación, es difícil estar seguro de que los controles, como quiera que se elijan, son realmente similares en todos los demás aspectos salvo en la exposición, lo que es un elemento clave para la validez de este tipo de estudios (v. el siguiente apartado). Por fortuna, en la actualidad, pocas veces se da la circunstancia de tener que asumir semejante riesgo científico, puesto que existen numerosas bases de datos que permiten

booksmedicos.org

muestreos de población real.

Independientemente de cómo se identifiquen los casos, deben poder quedar expuestos al factor de riesgo y deben poder presentar el resultado. Por ejemplo, en un estudio de casos y controles y la muerte súbita, los casos y los controles tenían que poder practicar un grado similar de ejercicio para ser elegibles.

Huelga decir que el diagnóstico debe ser rigurosamente confirmado en los casos (y excluido en los controles), y los criterios deben quedar bien explícitos. En el estudio sobre los bisfosfonatos, los investigadores estuvieron de acuerdo sobre los criterios explícitos de las fracturas atípicas del fémur y revisaron todas las radiografías, no sólo los informes de estas, para clasificar el tipo de fractura. Luego, un investigador revisó una muestra aleatoria de radiografías por segunda ocasión sin saber cómo se había clasificado cada una, y la primera y la segunda clasificación coincidieron completamente.

Selección de controles

La validez de los estudios de casos y controles depende, por encima de todo, de la posibilidad de comparación entre casos y controles. Para ser comparables, los casos y los controles deben pertenecer a la misma población inicial y tener la misma oportunidad de estar expuestos. El mejor método para cumplir con estos requerimientos es asegurarse que los controles son una muestra aleatoria de aquellos que no son casos en la misma población o cohorte de donde se extrajeron los casos.

El método poblacional

Los estudios en los que casos y controles constituyen una muestra completa o probabilística de una población definida se denominan **estudios de casos y controles basados en la población.** En la práctica, la mayoría de estas poblaciones son dinámicas, es decir, cambian continuamente, pues las personas entran y salen de la población, como se describió en el capítulo 2 (3). Esto puede sesgar los resultados, especialmente si la muestra de los casos y los controles se obtiene durante un largo período de tiempo y la exposición está cambiando rápidamente durante este tiempo. Esta preocupación puede resolverse si existe evidencia que la rotación en la población es de hecho tan baja que tiene poco efecto sobre los resultados del estudio o si los casos y controles están pareados de acuerdo con el calendario, es decir, se seleccionan los controles la misma fecha en que apareció la enfermedad en

booksmedicos.org

los casos.

El método de cohorte

Otra forma de garantizar que casos y controles son comparables es extraerlos de la misma cohorte. En esta situación, el estudio recibe el nombre de **estudio de casos y controles anidado** (puesto que está «anidado» en la cohorte).

En la era de las grandes bases de datos y las computadoras de gran capacidad, ¿por qué no simplemente analizar los datos de una cohorte como un estudio de cohorte en lugar de un estudio de casos y controles? Después de todo, la ineficacia de incluir muchos miembros expuestos de la cohorte, a pesar de que pocos de ellos presentarán el resultado, podría ser superada por el poder informático. La razón habitual para analizar los datos de una cohorte como un estudio de casos y controles es que algunas de las variables del estudio pueden no estar disponibles en la base de datos de la cohorte y, por lo tanto, deben recopilarse de otras fuentes para cada paciente del estudio. Puede ser muy costoso y lento recopilar la información que falta de registros médicos, cuestionarios, análisis genéticos y vincularla con otras bases de datos. Por lo tanto, existe una ventaja práctica de reunir esta información sólo para los casos y para una muestra de controles en la cohorte y no para cada miembro de la cohorte.

Con los estudios de casos y controles anidados existe una oportunidad de obtener medidas crudas de la incidencia de un análisis de cohortes y una fuerte estimación del riesgo relativo, que tiene en cuenta un amplio conjunto de covariables a partir de un análisis de casos y controles. Con esta información se tiene el conjunto completo del riesgo descrito en el capítulo 5, el riesgo absoluto en las personas expuestas y no expuestas, el riesgo relativo, y el riesgo atribuible poblacional.

El ejemplo de los bisfosfonatos muestra las ventajas de una cohorte complementaria y un análisis de casos y controles. Un análisis de cohorte que tiene en cuenta solamente la edad, mostró que el aumento del riesgo absoluto de fracturas atípicas relacionado con los bisfosfonatos fue de cinco casos por 10 000 años-paciente. La recopilación de los datos de las covariables se realizó mediante la vinculación con otras bases de datos y probablemente requería de muchos recursos para realizarlo en toda la muestra nacional. Con estos datos para los casos y controles fue posible obtener una estimación del riesgo relativo mucho más creíble con el análisis de casos y controles. La estimación de un riesgo relativo de 33, obtenido del análisis de casos y

booksmedicos.org

controles, fue consistente con el riesgo relativo crudo de 47, obtenido con el análisis de cohortes (que no consideró factores potenciales de confusión además de la edad). Debido a ambos análisis, los autores pudieron señalar que el riesgo relativo de fractura atípica era alto, pero el riesgo absoluto era bajo.

Controles hospitalarios y de la comunidad

Si no se puede realizar un muestreo basado en la población o en cohortes, una alternativa sería seleccionar los controles de forma que sean, *aparentemente,* comparables con los casos. Por ejemplo, si los casos seleccionados proceden de un centro hospitalario, los controles deberían seleccionarse entre pacientes con diferentes enfermedades, que aparentemente no guardan relación con la exposición ni con la enfermedad objeto de interés en el mismo hospital. Como se comentó antes, para la mayoría de los factores de riesgo y enfermedades, los estudios de casos y controles en los centros sanitarios tienen más posibilidad de fallar que en la población o basado en cohortes, debido a que los pacientes hospitalizados suelen constituir una muestra sesgada de todas las personas de la comunidad, las personas a quienes deben aplicarse los resultados.

Otro método consiste en obtener los controles de la comunidad a la que presta asistencia el hospital. Sin embargo, muchos hospitales no atienden de forma exclusiva a los pacientes de la comunidad a la que pertenecen, sino que muchas personas de la misma se dirigen a otros hospitales, y personas de otras comunidades prefieren recibir asistencia en el hospital en el que se realiza el estudio, en lugar de en su hospital más próximo. Como consecuencia, puede que casos y controles difieran de forma sustancial en aspectos que distorsionan la relación entre la exposición y la enfermedad.

Múltiples grupos de control

Si ninguno de los grupos de control disponibles parece idóneo, es posible observar cómo la elección de los controles afecta a los resultados, seleccionando varios grupos de control con puntos científicamente fuertes y puntos científicamente débiles, pero aparentemente complementarios. Si se obtienen estimaciones de riesgo relativo similares usando diferentes grupos de control, esto es una prueba de que no han producido sesgos, puesto que es improbable que los mismos sesgos afecten en la misma dirección y en la misma medida a grupos que, por lo demás, son diferentes. Si las estimaciones

booksmedicos.org

de riesgo relativo son distintas, esto indica que uno o ambos grupos están sesgados, y que deben investigarse las razones.

Ejemplo

En el ejemplo sobre el uso de cascos y traumatismo craneoencefálico (2), el grupo principal de control fueron las personas sin lesiones que esquiaban o practicaban snowboard en las mismas colinas los mismos días, pero podemos imaginarnos las desventajas de estos controles, como no tener una conducta de riesgo similar a los casos. Para examinar el efecto de la elección del grupo control sobre los resultados, los investigadores repitieron los análisis con un grupo control diferente: esquiadores con otras lesiones. La estimación del riesgo relativo fue similar, una reducción del riesgo de 55 % en lugar del 60 % con el grupo control original, lo que sugiere que la elección del grupo control no afectó sustancialmente a los resultados.

Múltiples controles por caso

El hecho de contar con varios *grupos* de control por grupo de casos no debe confundirse con el hecho de contar con varios controles por caso. Si el número de casos es limitado, como suele suceder con las enfermedades raras, el estudio puede aportar más información cuando hay más de un control por caso. Un mayor número de controles aumenta la capacidad de detectar un incremento o una reducción del riesgo, en caso de que este exista, una propiedad del estudio denominada «potencia estadística» (v. cap. 11). Desde el punto de vista práctico, vale la pena ganar en capacidad hasta que se disponga de tres o cuatro controles por caso, tras lo cual, un mayor número de controles aportaría muy poca mejoría adicional.

Pareamiento

Si algunas características parecen guardar una relación especialmente estrecha con la exposición o la enfermedad, de modo que se deseara estar seguro que se producen de manera similar en los casos y en los controles, podrían parearse. Mediante el **pareamiento,** para cada caso con un conjunto de características, se seleccionan uno o más controles con las mismas características. Los investigadores suelen realizar pareamientos por edad y sexo porque a menudo estos aspectos están fuertemente relacionados con la exposición y con la enfermedad, pero el pareamiento suele ampliarse más allá de estas características demográficas (p. ej., con el perfil de riesgo o la gravedad de la enfermedad) cuando se sabe que otros factores están

booksmedicos.org

estrechamente relacionados con la exposición o el resultado. El pareamiento aumenta la información útil que puede obtenerse de un conjunto de casos y controles, ya que reduce las diferencias entre los grupos en cuanto a factores determinantes de una enfermedad aparte del factor estudiado, lo que aumenta la potencia (sensibilidad) de la prueba de asociación.

En ocasiones, los casos y controles son comparables con el **pareamiento en sombrilla,** basado en una variable como el hospital o la comunidad, que es representativa de muchas otras variables que podrían confundir la relación entre la exposición y la enfermedad, y serían difíciles de medir una cada vez, si esto fuera posible. Los ejemplos de las variables que podrían ser contempladas bajo una sombrilla incluyen la desigualdad social relacionada con el ingreso, la educación, la raza y la etnia; la tendencia a buscar atención médica o seguir el consejo médico, y los patrones locales para la atención médica.

El pareamiento puede ser exagerado, sesgando los resultados del estudio. El **sobrepareamiento** puede ocurrir si los investigadores parean variables muy íntimamente relacionadas con la exposición que las tasas de exposición en los casos y controles se vuelven más similares de lo que son en la población. El resultado es hacer que la estimación observada del riesgo relativo se acerque a 1 (sin efecto). Hay muchas razones por las que la variable pareada podría estar relacionada con la exposición. Podría ser parte de la cadena de acontecimientos que llevan de la exposición a la enfermedad. Otras variables podrían estar relacionadas entre sí porque tienen causas fundamentales similares; la educación, los ingresos, la raza y la etnia tienden a estar relacionadas entre sí, por lo que si pareamos de acuerdo con una de ellas, se enmascararán los efectos de las otras. El pareamiento de enfermedades con el mismo tratamiento podría causar sobrepareamiento en los estudios sobre los efectos de ese tratamiento. Por ejemplo, en un estudio sobre los anti-inflamatorios no esteroideos (AINE) y la insuficiencia renal, si los casos y controles fueran pareados por la presencia de síntomas artríticos, que a menudo se tratan con AINE, los pares pareados podrían tener el mismo antecedente, aunque artificial, sobre el uso de AINE.

Una desventaja del pareamiento es que una vez que se parean según una variable, se vuelve similar en los casos y en los controles, y ya no es posible conocer cómo afecta a la relación entre la exposición y la enfermedad. También, para muchos estudios no es posible encontrar controles pareados para más de unas cuantas características del caso. Esto puede superarse, en

booksmedicos.org

cierta medida, si el número de posibles controles es enorme o si los criterios de pareamiento son holgados (p. ej., parear la edad con un intervalo de 5 años en lugar de la misma edad). En resumen, el pareamiento es una manera útil de controlar la confusión, pero puede limitar las preguntas formuladas en el estudio y causar sesgos en lugar de eliminarlos.

Medida de la exposición

La validez de los estudios de casos y controles también depende de si se evita una clasificación incorrecta cuando se mide la exposición. El método más seguro para medir la exposición consiste en acudir a registros exactos y completos recopilados antes de que se produjera la enfermedad. Por ejemplo, los registros farmacéuticos en estudios sobre riesgos en la prescripción de un fármaco, los registros quirúrgicos en estudios sobre complicaciones quirúrgicas y las muestras de sangre almacenadas para estudios sobre riesgo relacionado con anomalías bioquímicas. Con este tipo de registros, el conocimiento del estado de la enfermedad no puede sesgar la notificación de la exposición. Sin embargo, muchas exposiciones importantes sólo pueden medirse preguntando a casos y controles, o a sus allegados. Entre ellas se encuentran el ejercicio, la alimentación y el consumo de fármacos de libre dispensación o de drogas. El siguiente ejemplo ilustra cómo los investigadores pueden «fortalecer» los datos de las entrevistas que son inherentemente vulnerables al sesgo.

Ejemplo

¿Cuáles son los factores de riesgo de suicidio en China? Los investigadores estudiaron 519 personas que se habían suicidado y 536 que habían fallecido por otros daños físicos (4); ambos grupos procedían de 23 centros geográficamente representativos de China. La exposición se midió mediante entrevistas con familiares y otros allegados. Los autores señalaron que, al igual que otros estudios que dependían de la «autopsia psicológica» para medir la exposición, «los entrevistadores conocían la causa de la muerte del fallecido (suicidio u otros daños físicos), por lo que no pudimos eliminar por completo el posible sesgo del entrevistador». Los autores siguieron explicando que «intentamos reducir este sesgo al mínimo utilizando el mismo esquema de entrevista para casos y controles, empleando medidas objetivas de posibles factores de riesgo, obteniendo indicios de forma independiente a partir de dos fuentes (familiares y allegados) y proporcionando una amplia formación a los entrevistadores». Asimismo, seleccionaron a controles fallecidos por daños físicos que pudieran parearse en función de una importante característica que podría haber afectado a las respuestas proporcionadas en las entrevistas. Dicha característica fue la muerte reciente de un familiar o de un allegado.

booksmedicos.org

El estudio identificó ocho predictores de suicidio: puntuación elevada de síntomas depresivos, intento de suicidio previo, estrés agudo previo a la muerte, escasa calidad de vida, estrés crónico elevado, conflicto interpersonal grave en los 2 días anteriores a la muerte, pariente consanguíneo con conducta suicida previa, y amigo o allegado con conducta suicida previa.

Cuando se pide a casos y controles que recuerden sus exposiciones previas, pueden presentarse sesgos por diferentes razones. Es probable que los casos, al saber que su enfermedad está siendo estudiada, puedan ser más propensos a recordar si han estado expuestos, un problema denominado **sesgo de memoria.** Por ejemplo, los padres de un niño con síndrome de Reye (una encefalopatía) pueden ser más propensos a recordar el uso de aspirina después de los esfuerzos generalizados para que los padres estén al tanto de una asociación entre el uso de aspirina y el síndrome de Reye en los niños con fiebre. Un hombre con cáncer de próstata tiene más probabilidades de notificar una vasectomía previa después de que las historias de asociación aparecieron en las noticias. Con toda la publicidad que rodea a los posibles riesgos de las diferentes exposiciones farmacológicas y ambientales, es perfectamente posible que las víctimas de la enfermedad recuerden sus exposiciones más a menudo que las personas sin ella.

Los investigadores pueden limitar el sesgo de memoria al no informar a los pacientes sobre el propósito específico del estudio. No sería ético no informar a los participantes en la investigación de la naturaleza general del objeto del estudio, pero proporcionarles información detallada sobre aspectos e hipótesis concretos podría sesgar de tal manera la información resultante que se incumpliría otro principio ético: implicar a personas en un proyecto de investigación que no tendrá utilidad alguna.

Los médicos pueden ser más propensos a preguntar sobre una exposición y registrar esta información en el informe médico en los casos que en los controles si se sospecha previamente que la exposición es una causa. Por lo tanto, es más probable que un médico registre el antecedente familiar de cáncer de próstata en un paciente con cáncer de próstata o que registre el uso de teléfono móvil en un paciente con cáncer cerebral. Este sesgo debería quedar claro para todos los que estudian el diagnóstico clínico. Si un médico residente que atiende a una mujer relativamente joven con un infarto agudo de miocardio conoce la asociación notificada con el uso de anticonceptivos orales, es probable que insista más en preguntar a la paciente por el uso de anticonceptivos orales, y que registre esta información de forma más minuciosa. Las protecciones frente a este tipo de sesgo son las mismas que

booksmedicos.org

las mencionadas anteriormente: diversas fuentes de información y el cegamiento de las personas encargadas de obtener los datos, evitando que conozcan las hipótesis en estudio.

La exposición puede verse afectada por la presencia de la enfermedad, especialmente cuando la exposición objeto del estudio es un tratamiento farmacológico. Las manifestaciones iniciales de la enfermedad pueden conllevar un tratamiento, mientras que la pregunta del estudio es justo lo opuesto: averiguar si el tratamiento causa enfermedad. Si se anticipa este problema, existe la posibilidad de atajarlo, tal como ilustra el ejemplo siguiente.

Todo lo que puede decirse sobre el sesgo en la medición de la exposición también se puede decir sobre los factores de confusión. Muchas covariables importantes (p. ej., tabaquismo, dieta, ejercicio, así como la raza y la etnia) pueden ser registradas inadecuadamente en los registros y las bases de datos médicos y, por lo tanto, deben ser obtenidos mediante entrevistas para que puedan incluirse en el estudio.

Ejemplo

¿Previenen los β-bloqueadores los primeros infartos de miocardio en pacientes que reciben tratamiento para la hipertensión arterial? Para responder a esta pregunta se realizó un estudio de casos y controles (5). Puesto que la angina de pecho es una importante indicación del uso de fármacos β-bloqueadores, y también un síntoma de enfermedad coronaria, los investigadores excluyeron minuciosamente a todos los pacientes con antecedentes que sugerían la presencia de angina de pecho o con antecedentes de cualquier otra manifestación de cardiopatía coronaria. Observaron que los pacientes con hipertensión arterial tratados con β-bloqueadores presentaban un riesgo reducido de infartos de miocardio no mortales, incluso después de excluir de forma minuciosa a los pacientes con angina de pecho o con cualquier otro indicio de enfermedad coronaria.

Exposiciones múltiples

Hasta el momento hemos descrito estudios de casos y controles de una sola exposición dicotómica, pero los estudios de casos y controles son un medio eficaz para examinar una variedad mucho más rica de exposiciones: los efectos de múltiples exposiciones, varias dosis de la misma exposición y las exposiciones que son síntomas tempranos (no factores de riesgo) de la enfermedad.

booksmedicos.org

Ejemplo

El cáncer de ovario es muy difícil de diagnosticar al inicio de su curso, cuando el tratamiento podría ser más efectivo. En Inglaterra, los investigadores realizaron un estudio de casos y controles sobre los síntomas del cáncer de ovario en la atención primaria (6). Los casos fueron 212 mujeres mayores de 40 años diagnosticadas con cáncer de ovario primario en 39 consultas generales en Devon, Inglaterra, de 2000 a 2007; los casos se parean con 1 060 controles con cáncer de ovario por edad y consulta. Los síntomas se obtuvieron de los registros médicos del año antes del diagnóstico. Siete síntomas se asociaron independientemente con el cáncer de ovario: distensión abdominal, sangrado postmenopáusico, pérdida de apetito, aumento de la frecuencia urinaria, dolor abdominal, sangrado rectal e inflamación abdominal. Después de excluir los síntomas aportados los 180 días previos al diagnóstico (para tener una mejor estimación de los síntomas «tempranos»), tres de ellos continuaron asociados de forma independiente con el cáncer de ovario: la distensión abdominal, la frecuencia urinaria y el dolor abdominal.

RAZÓN DE MOMIOS *(ODDS RATIO):* UNA ESTIMACIÓN DEL RIESGO RELATIVO

La figura 6-3 muestra la clasificación dicotómica de la exposición y la enfermedad, típica tanto de los estudios de cohorte como los estudios de casos y controles, y compara cómo se calcula el riesgo de forma diferente para los dos. Estos conceptos se ilustran con el estudio de los bisfosfonatos, que tiene un componente tanto de cohortes como de casos y controles.

En el estudio de cohorte, los participantes se dividieron en dos grupos al inicio expuestos a bisfosfonatos ($a + b$) y no expuestos a bisfosfonatos ($c + d$). Los casos de fracturas atípicas se presentaron naturalmente durante el tiempo en el grupo expuesto (a) y en el grupo no expuesto (c), lo que proporciona numeradores y denominadores apropiados para calcular la incidencia de fracturas atípicas en los miembros expuestos $[a/(a + b)]$ y no expuestos $[c/(c + d)]$ de la cohorte. También fue posible calcular el riesgo relativo:

booksmedicos.org

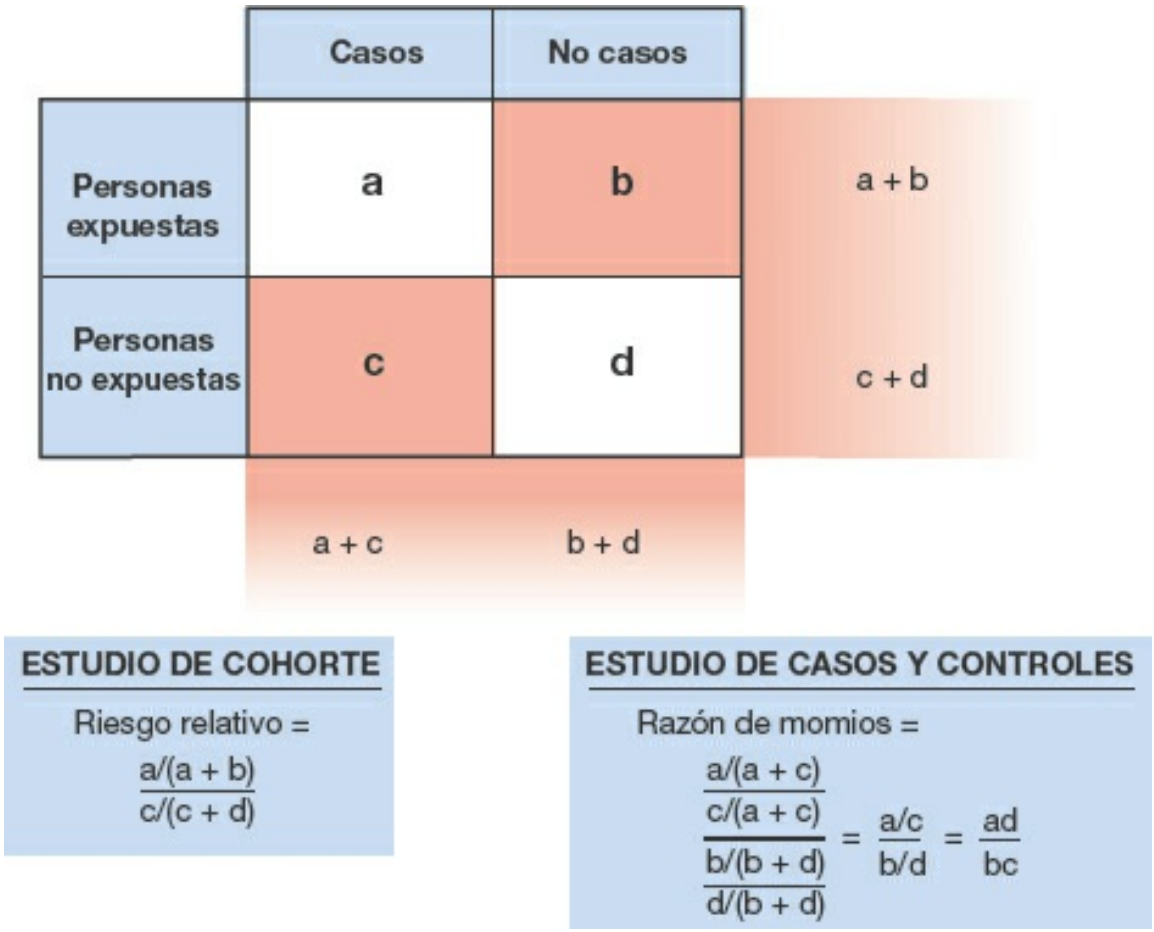


Figura 6-3 ■ Cálculo del riesgo relativo en un estudio de cohorte y de la razón de momios (riesgo relativo estimado) en un estudio de casos y controles.

Riesgo relativo

$$= \frac{\text{Incidencia de enfermedad en la cohorte expuesta}}{\text{Incidencia de enfermedad en la cohorte no expuesta}}$$

$$= \frac{a/(a+b)}{c/(c+d)}$$

Por otro lado, el estudio de casos y controles comenzó con la selección de un grupo de casos de fractura atípica ($a + c$) y un grupo de controles con otras fracturas ($b + d$). No hay forma de conocer las tasas de enfermedad, ya que estos grupos no se determinan de forma natural sino en función de los criterios de selección del investigador. Por ello, no es posible calcular la incidencia de la enfermedad entre las personas expuestas y no expuestas a los bisfosfonatos; y no es posible calcular el riesgo relativo. Sin embargo, lo que cobra importancia es la frecuencia relativa de la exposición a los

booksmedicos.org

bisfosfonatos en los casos y los controles.

Un método para comparar la frecuencia de exposición entre casos y controles proporciona una medida de riesgo que es, conceptual y matemáticamente, similar al riesgo relativo. Es lo que se conoce como **razón de momios *(odds ratio),*** que se define como el momio (*odds*) de que un caso esté expuesto divididas por el momio de que un control esté expuesto:

$$\frac{\left(\frac{a}{a+c}\right)}{\left(\frac{c}{a+c}\right)}$$
$$\frac{\left(\frac{b}{b+d}\right)}{\left(\frac{d}{b+d}\right)}$$

La razón de momios se simplifica a:

$$\frac{\left(\frac{a}{c}\right)}{\left(\frac{b}{d}\right)},$$

donde el momio es el cociente de dos probabilidades, la probabilidad de un evento dividido entre 1 – la probabilidad del mismo acontecimiento.

La razón de momios se puede simplificar aún más de la siguiente manera:

$$\frac{ad}{bc}$$

Como se observa en la figura 6-3, la razón de momios se puede obtener multiplicando las cifras de la tabla en diagonal y dividiendo estos productos cruzados.

El significado de la razón de momios es análogo al riesgo relativo obtenido en los estudios de cohorte. Si la frecuencia de exposición es más elevada entre los casos, la razón de momios será superior a 1, lo que indicará un mayor riesgo. Por consiguiente, cuanto más estrecha sea la asociación

booksmedicos.org

entre la exposición y la enfermedad, más alta será la razón de momios. Por el contrario, si la frecuencia de exposición es más reducida entre los casos, la razón de momios será inferior a 1, lo que indicará protección. Debido a la similitud entre la información expresada por la razón de momios y la expresada por el riesgo relativo, y el significado más fácilmente vinculado al riesgo relativo, algunos investigadores notifican las razones de momios como **riesgos relativos estimados.**

La razón de momios es aproximadamente equivalente al riesgo relativo sólo cuando la incidencia de la enfermedad es reducida. Para su representación matemática, observe la fórmula para el riesgo relativo en la figura 6-3. Si el número de casos en el grupo expuesto (*a*) es pequeño en relación con el número de controles en ese grupo (*b*), entonces $a/(a + b)$ es aproximadamente igual a a/b. De forma similar, si el número de casos en el grupo no expuesto (*c*) es pequeño en relación con el número de controles en ese grupo (*d*), entonces $c/(c + d)$ se calcula como c/d. Así, el riesgo relativo = a/b dividido entre c/d, que puede simplificarse como ad/bc, la razón de momios.

¿Qué tan bajas deben ser las tasas para que la razón de momios sea una estimación exacta del riesgo relativo? La respuesta depende del tamaño del riesgo relativo (7). En general, los sesgos de la estimación del riesgo relativo se vuelven lo suficientemente importantes como para tener un significado en función de las tasas de enfermedad en las personas no expuestas para ser mayores de aproximadamente 1/100 o quizás 5/100. Conforme los resultados se vuelven cada vez más frecuentes, la razón de momios tiende a sobreestimar el riesgo relativo cuando es > 1 y a subestimarlo cuando el riesgo relativo es < 1. Por fortuna, la mayoría de las enfermedades, en particular las estudiadas a través de estudios de casos y controles, tienen tasas considerablemente más bajas.

Al inicio de este capítulo describimos la razón por la que los estudios de casos y controles deben ser sobre casos incidentes (comienzo reciente), no sobre casos prevalentes. Sin embargo, las **razones de momios para la prevalencia** se calculan a menudo en los estudios de prevalencia y se reportan en la literatura médica. La razón de momios de prevalencia es una medida de asociación pero no es muy informativa, no sólo debido a los factores relacionados con la dificultad para distinguir la incidencia frente a la duración, sino también porque es menos probable que se cumpla el supuesto de una enfermedad rara.

booksmedicos.org

CONTROL DE LAS VARIABLES EXTRAÑAS

La mayor amenaza de la validez de los estudios observacionales (cohortes y de casos y controles) es que los grupos que se comparan pueden ser sistemáticamente diferentes en factores relacionados con la exposición y la enfermedad, es decir, hay confusión. En el capítulo 5 describimos varias formas de controlar las variables extrañas cuando buscamos los efectos independientes de la exposición sobre la enfermedad en los estudios observacionales. Todos estos métodos, exclusión, pareamiento, análisis estratificado y modelado, también se utilizan en los estudios de casos y controles, a menudo en combinación. Obviamente, esto sólo puede realizarse para aquellas características que ya se sospechaba que afectaban a la relación entre exposición y enfermedad, y que se midieron en el estudio.

Ya que casi siempre se usa el modelado matemático para controlar las variables extrañas, en la práctica, el cálculo de la razón de momios es mucho más complicado que el producto cruzado de una tabla de dos por dos. Una razón de momios calculada directamente de una tabla de 2 x 2 se conoce como una **razón de momios cruda,** porque no ha tenido en cuenta otras variables además de la exposición y la enfermedad. Después de ajustar los efectos de estas otras variables, se denomina **razón de momios ajustada.**

La razón implícita de los estudios de casos y controles es hallar las causas. Sin embargo, incluso aunque se hayan controlado las variables extrañas mediante métodos científicos, persiste la posibilidad de que las variables no medidas confundan la relación entre la exposición y la enfermedad. Por lo tanto, uno debe conformarse con la descripción de cómo la exposición está relacionada con la enfermedad, independientemente de otras variables incluidas en el estudio y considerar humildemente la posibilidad de que las variables no medidas puedan explicar los resultados. Por estas razones, los resultados de los estudios observacionales se describen mejor como asociaciones, no como causas.

INVESTIGACIÓN DE UN BROTE EPIDÉMICO

Hasta el momento, hemos descrito el uso del método de casos y controles para identificar factores de riesgo de enfermedades crónicas. El mismo método se utiliza para identificar factores de riesgo de brotes (pequeñas epidemias) de enfermedades agudas, habitualmente infecciosas o intoxicaciones. Con frecuencia, la causa microbiológica de la enfermedad

booksmedicos.org

resulta obvia al principio de la epidemia, tras la evaluación diagnóstica de los casos, pero puede que el modo de transmisión no sea tan evidente, y se precise información sobre la forma de propagación de la enfermedad para detener la epidemia. Esta información también incrementa el conocimiento sobre los posibles modos de transmisión, lo que puede ser útil para el control de futuras epidemias.

Ejemplo

Un gran brote de gastroenteritis, con muchos casos complicados de síndrome urémico hemolítico (un trastorno potencialmente mortal con insuficiencia renal aguda, anemia hemolítica y trombocitopenia) se presentó en Alemania en mayo de 2011 (8). Durante esta epidemia, se informaron de 3 816 casos, 845 con síndrome urémico hemolítico. La figura 6-4 muestra la **curva epidémica,** el número de casos a lo largo del tiempo. La causa inmediata se identificó rápidamente, infección por una cepa productora de toxinas de la bacteria *Escherichia coli*, no así la fuente de infección. Los investigadores realizaron un estudio de casos y controles comparando 26 casos de síndrome urémico hemolítico con 81 controles, pareados por edad y vecindario (9). Encontraron que 6/24 casos (25 %) y 7/80 controles (9 %) estuvieron expuestos a las coles, con una razón de momios de 5,8, lo que sugiere que la infección fue transmitida por comer coles contaminadas. (Observe que para este caso la razón de momios no son exactamente los productos cruzados porque su cálculo tiene en cuenta el pareamiento). Sin embargo, los pepinos y otros productos agrícolas también estuvieron implicados, aunque con menos fuerza. Para saber más, los investigadores llevaron a cabo un pequeño estudio de cohorte en personas que cenaban en grupos en un solo restaurante durante el período de la epidemia. Se definió empíricamente a los casos como los comensales que presentaron diarrea sanguinolenta o síndrome urémico hemolítico, o bien un cultivo con el microorganismo causal. El 20 % de la cohorte cumplió con estos criterios, el 26 % de estos tenía síndrome urémico hemolítico. El riesgo relativo para el consumo de coles fue 14,2 y ningún otro alimento se asoció fuertemente con la enfermedad. El consumo de coles representó el 100 % de los casos. Los investigadores rastrearon el origen de las coles desde el distribuidor que abasteció al restaurante hasta un solo productor; sin embargo, no lograron cultivar *E. coli* causal de las semillas en el lote implicado. Después de la investigación y de haber llamado la atención al productor, la epidemia disminuyó (fig. 6-4) y la incidencia volvió a los niveles bajos observados antes de la epidemia.

Este ejemplo también ilustra cómo los estudios de casos y controles y los de cohorte, los estudios de laboratorio para buscar el microorganismo causal y la epidemiología de campo (de intervención) durante el rastreo actuaron en conjunto para identificar la causa subyacente de la epidemia.

booksmedicos.org

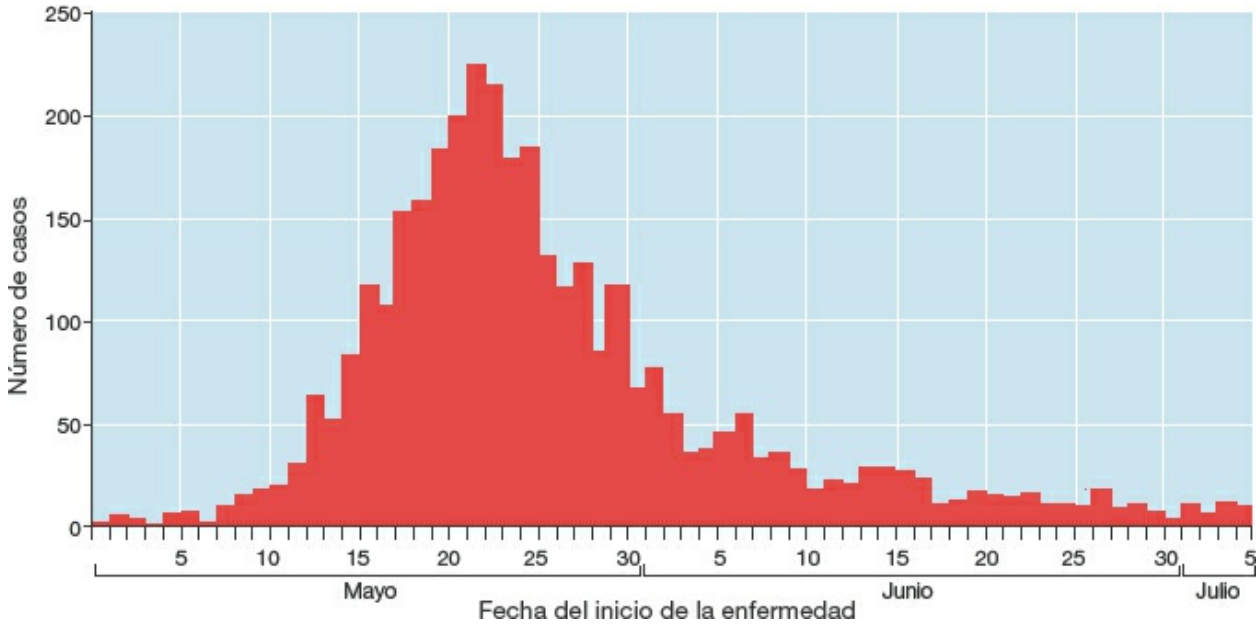


Figura 6-4 Curva epidémica de un brote de infección por *Escherichia coli* productora de toxina Shiga en Alemania. (Redibujado con autorización de Frank C, Werber D, Cramer JP, et al. Epidemic profile of shiga-toxin-producing Escherichia coli 0104:H4 outbreak in Germany. N Engl J Med 2011;365:1771-1780.)

Preguntas de revisión

Lea los siguientes enunciados y seleccione la mejor respuesta.

6-1. En un estudio de casos y controles sobre anti-conceptivos orales e infarto agudo de miocardio se determinó la exposición a las pastillas anticonceptivas de registros médicos en el momento del infarto agudo de miocardio. Los resultados podrían estar sesgados para buscar una asociación a través de todos los factores siguientes, *excepto:*

A. Los médicos podrían haber interrogado más cuidadosamente a los casos sobre el uso de anticonceptivos orales.

B. Haber tenido un infarto de miocardio podría dar lugar al uso de anticonceptivos orales.

C. Es probable que los médicos hayan tenido más posibilidades de registrar el uso de anti-conceptivos orales en los casos.

D. Los revisores de los registros médicos podrían haber buscado la evidencia sobre el uso de anticonceptivos orales más cuidadosamente

booksmedicos.org

si supieran que un paciente había tenido un infarto de miocardio.

E. Los pacientes podrían haber recordado la exposición más fácilmente cuando tuvieron un ataque cardíaco.

6-2. Investigadores europeos llevaron a cabo un estudio de casos y controles anidado en una cohorte, proveniente de varios países con más de 520 000 participantes, sobre la concentración de vitamina D y el riesgo de cáncer de colon. Estudiaron 1 248 casos de cáncer de colon incidental que surgieron en la cohorte y un número igual de controles, obtenidos de la misma cohorte y pareados por edad, sexo y centro de estudio. La vitamina D se midió en muestras de sangre tomadas años antes del diagnóstico. Los niveles de vitamina D fueron menores en pacientes con cáncer de colon, independientemente de una serie de posibles variables confusoras. Los resultados del estudio podrían explicarse por cualquiera de los siguientes, *excepto:*

A. Los niveles de vitamina D se asocian con cáncer colorrectal.

B. La deficiencia de vitamina D es un factor de riesgo para el cáncer colorrectal.

C. Anidar el estudio en una cohorte numerosa fue una de sus fortalezas.

D. Los resultados podrían haberse confundido por variables no medidas relacionadas con los niveles de vitamina D y el cáncer colorrectal.

E. La deficiencia de vitamina D fue la causa del cáncer colorrectal.

6-3. ¿Cuál de los siguientes es el resultado más directo de un estudio de casos y controles?

A. Prevalencia.

B. Diferencia del riesgo.

C. Riesgo relativo.

D. Incidencia.

E. Razón de momios.

6-4. La curva epidémica para una enfermedad infecciosa aguda describe:

A. El período de incubación habitual para el agente causal.

booksmedicos.org

B. La comparación de la enfermedad a lo largo del tiempo en las personas expuestas y no expuestas.

C. El inicio de la enfermedad en los casos a lo largo del tiempo.

D. La duración promedio de la enfermedad en los individuos afectados.

E. La distribución del tiempo de la infección hasta la aparición de los primeros síntomas.

6-5. ¿Cuál de los siguientes es el mejor motivo para hacer un análisis de casos y controles en un estudio de cohorte?

A. Los estudios de casos y controles son una forma factible de controlar los factores de confusión que no se encuentran en la base de datos de la cohorte.

B. Los estudios de casos y controles pueden proporcionar la misma información con mayor facilidad.

C. Los estudios de casos y controles pueden determinar la incidencia de la enfermedad en los miembros expuestos y no expuestos de la cohorte.

D. Los estudios de casos y controles son en general más fuertes que los estudios de cohorte.

6-6. La mejor forma para identificar los casos es obtenerlos a partir de:

A. Una muestra de la población general (dinámica).

B. Consultas de médicos de atención primaria.

C. Una comunidad.

D. Una cohorte representativa de la población.

E. Un hospital.

6-7. ¿Cuál es el mejor motivo para incluir múltiples grupos de control en un estudio de casos y controles?

A. Para obtener una estimación más fuerte del riesgo relativo.

B. Existe número limitado de casos y un gran número de posibles controles.

C. Para controlar la confusión.

D. Para aumentar la generalización del resultado.

booksmedicos.org

E. El grupo control principal puede ser sistemáticamente diferente de los casos (distinto a la exposición de interés).

6-8. Los estudios de casos y controles se pueden utilizar para estudiar todo lo siguiente, *excepto:*

A. Los síntomas precoces del cáncer de estómago.
B. Los factores de riesgo del síndrome de muerte súbita infantil.
C. La incidencia de suicidio en la población adulta.
D. El efecto protector de la aspirina.
E. Las formas de transmisión de una enfermedad infecciosa.

6-9. En un estudio de casos y controles sobre el ejercicio y la muerte súbita, ¿para qué podría ser útil el pareamiento?

A. Para controlar todas las posibles variables confusoras en el estudio.
B. Para hacer que algunas de las características principales de los casos y controles sean similares entre sí.
C. Para que sea posible examinar los efectos de las variables pareadas en la estimación del riesgo relativo.
D. Para probar si se eligieron los controles correctos para los casos en el estudio.
E. Para aumentar la generalización del estudio.

6-10. En un estudio de casos y controles para determinar si el viaje aéreo prolongado es un factor de riesgo para tromboembolia venosa, 60 de cada 100 casos y 40 de cada 100 controles tuvieron viajes aéreos prolongados. ¿Cuál fue la razón de momios cruda para este estudio?

A. 0,44.
B. 1,5.
C. 2,25.
D. 3,0.
E. No es posible calcularlo.

6-11. Un estudio de casos y controles basado en una población sería especialmente útil para estudiar:

A. El riesgo atribuible poblacional de una enfermedad.

booksmedicos.org

B. Múltiples resultados (enfermedades).
C. La incidencia de enfermedades infrecuentes.
D. La prevalencia de la enfermedad.
E. Los factores de riesgo de la enfermedad.

6-12. La razón de momios de prevalencia de la artritis reumatoide consiste en un estimado de:

A. El riesgo relativo de artritis.
B. El riesgo atribuible de artritis.
C. Los factores de riesgo para la duración de la artritis.
D. La asociación entre un paciente característico y la prevalencia de artritis.
E. Los factores de riesgo para la incidencia de artritis.

6-13. En un brote de gastroenteritis aguda, un estudio de casos y controles sería especialmente útil para identificar:

A. Las características de las personas afectadas.
B. El número de personas afectadas a lo largo del tiempo.
C. El microorganismo o toxina que causó el brote.
D. La forma de transmisión.
E. Dónde se originó el agente causal.

6-14. ¿Cuál de los siguientes objetivos se logra con el muestreo de casos y controles de una población definida o una cohorte?

A. Es la única forma de incluir casos incidentales (nuevos) de la enfermedad.
B. Evita la necesidad de criterios de inclusión y exclusión.
C. Tiende a incluir casos y controles que son similares entre sí con excepción de la exposición.
D. Parear los casos y controles de acuerdo con las variables relevantes.
E. Garantiza que los resultados puedan generalizarse.

6-15. Los estudios de casos y controles podrían ser útiles para responder a todas las siguientes preguntas, *excepto:*

booksmedicos.org

A. ¿Los fármacos que disminuyen el colesterol previenen la enfermedad coronaria?

B. ¿Son más frecuentes las complicaciones en la colecistectomía con fibra óptica que en la cirugía convencional (abierta)?

C. ¿El consumo de alcohol es un factor de riesgo para el cáncer de mama?

D. ¿Con qué frecuencia se producen complicaciones después de la colecistectomía con fibra óptica?

E. ¿Qué tan efectivos son los antibióticos para la otitis media?

6-16. En un estudio de casos y controles sobre el viaje en avión y la tromboflebitis se debe cumplir con todas las siguientes condiciones para que la razón de momios sea una estimación razonable del riesgo relativo, *excepto:*

A. Los controles y casos fueron una muestra de la misma población.

B. Los casos y controles cumplieron con los mismos criterios de inclusión y exclusión.

C. Se controlaron otras variables que podrían haber estado relacionadas con el viaje en avión y la trombosis.

D. Los casos y controles tenían la misma susceptibilidad para desarrollar tromboflebitis (p. ej., tenían la misma movilidad, peso, traumatismos recientes, antecedente de TEP), además del viaje en avión.

E. La incidencia de tromboflebitis fue mayor a 5/100.

Encontrará las respuestas en el apéndice A.

BIBLIOGRAFÍA

1. Schilcher J, Michaelsson K, Aspenberg P. Bisphosphonate use and atypical fractures of the femoral head. *New Engl J Med* 2011;364:1728–1737.
2. Sulheim S, Holme I, Ekeland A, et al. Helmet use and risk of head injuries in alpine skiers and snowboarders. *JAMA* 2006;295:919–924.
3. Knol MJ, Vandenbroucke JP, Scott P, et al. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *Am J Epidemiol* 2008;168: 1073–1081.
4. Phillips MR, Yang G, Zhang Y, et al. Risk factors for suicide in China: a national case-control psychological autopsy study. *Lancet* 2002;360:1728–1736.
5. Psaty BM, Koepsell TD, LoGerfo JP, et al. Beta-blockers and primary prevention of coronary heart disease in patients with high blood pressure. *JAMA* 1989;261:2087–2094.

booksmedicos.org

6. Hamilton W, Peters TJ, Bankhead C, et al. Risk of ovarian cancer in women with symptoms in primary care: population-based case-control study. *BMJ* 2009;339:b2998. doi:10.1136/ bmj.b2998
7. Feinstein AR. The bias caused by high value of incidence for p1 in the odds ratio assumption that 1-p1 is approximately equal to 1. *J Chron Dis* 1986;39:485–487.
8. Frank C, Werber D, Cramer JP, et al. Epidemic profile of shiga-toxin-producing *Escherichia coli* 0104:H4 outbreak in Germany. *N Engl J Med* 2011;365:1771–1780.
9. Buchholz U, Bernard H, Werber D et al. German outbreak of *Escherichia coli* 0104:H4 associated with sprouts. *N Engl J Med* 2011;365:1763–1770.

booksmedicos.org

Capítulo 7

Pronóstico

Él, quien podía distinguir correctamente a los que sobrevivirían o morirían, así como los que tendrían la enfermedad durante un tiempo más largo o más corto, debe, según su conocimiento y atención, ser capaz de hacer una estimación de todos los síntomas y sopesar racionalmente su capacidad de comparación.

—Hipócrates
460-375 a.C.

PALABRAS CLAVE

Pronóstico
Factores pronósticos
Curso clínico
Evolución natural
Tiempo cero
Cohorte de inicio
Fase de migración
Acontecimiento
Análisis de supervivencia
Método de Kaplan-Meier
Análisis de tiempo hasta el acontecimiento
Pacientes censurados
Razón morbilidad *(hazard ratio)*
Serie de casos
Informe de caso
Reglas de predicción clínica
Conjunto de entrenamiento

booksmedicos.org

Conjunto de prueba
Validación
Estratificación pronóstica
Sesgo de muestreo
Sesgo de migración
Abandonos
Sesgo de medición
Análisis de sensibilidad
Análisis del mejor caso/el peor caso

Cuando las personas enferman, les asaltan muchas dudas sobre la forma en que les afectará su enfermedad. ¿Es peligrosa? ¿Puedo morir? ¿Será dolorosa? ¿Durante cuánto tiempo podré continuar con mis actividades habituales? ¿Se curará del todo? La mayoría de los pacientes y familiares desean saber qué pueden esperar, incluso en las situaciones en las que poco se puede hacer con respecto a la enfermedad.

El **pronóstico** es una predicción de la evolución de la enfermedad después de su inicio. Este capítulo revisa las formas en que puede describirse el curso de una enfermedad. La intención es que los lectores comprendan mejor lo que constituye un cometido difícil, pero indispensable: predecir el futuro de los pacientes de la forma más exacta posible. El objetivo es evitar la expresión de pronósticos con imprecisión cuando es innecesario, y con certeza cuando pueden inducir a error.

Tanto médicos como pacientes quieren conocer el curso general de la enfermedad, pero desean ir más allá y adaptar esta información a su situación particular en la medida de lo posible. Por ejemplo, a pesar de que el cáncer de ovario suele ser mortal a largo plazo, las mujeres que lo presentan pueden vivir desde unos pocos meses hasta muchos años, y desean saber cuál es su caso particular dentro de estas posibilidades.

Los estudios sobre pronóstico son similares a los estudios de cohortes sobre riesgo. Se agrupan los pacientes que tienen en común una enfermedad o afección concreta, se realiza el seguimiento del grupo a partir de ese momento y se miden los resultados clínicos. Se identifican las características asociadas a un resultado concreto de la enfermedad, denominadas **factores pronósticos,** que son análogos a los factores de riesgo, salvo por el hecho de que representan una parte diferente del espectro de la enfermedad: de la enfermedad a los resultados. Los estudios de casos y controles realizados con personas afectadas por la enfermedad que presentan y no presentan un mal resultado también pueden contribuir a calcular el riesgo relativo asociado a

booksmedicos.org

varios factores pronósticos, si bien no pueden proporcionar información sobre las tasas de resultados (v. cap. 6).

DIFERENCIAS ENTRE LOS FACTORES PRONÓSTICOS Y DE RIESGO

Los factores pronósticos y de riesgo difieren en varios aspectos.

Los pacientes son diferentes

Los estudios sobre factores de riesgo suelen realizarse con personas sanas, mientras que los estudios sobre factores pronóstico se realizan con personas enfermas.

Los resultados son diferentes

Para el riesgo, el acontecimiento que cuenta es el inicio de la enfermedad. Para el pronóstico, cuenta un intervalo de consecuencias de la enfermedad, entre las que se incluyen la muerte, las complicaciones, la incapacidad y el sufrimiento.

Las tasas son diferentes

Los factores de riesgo suelen hacer referencia a acontecimientos con probabilidades reducidas. Las tasas anuales del inicio de varias enfermedades son del orden de 1/1 000 a 1/100 000 o inferiores. Como consecuencia, es difícil confirmar las relaciones entre exposición y enfermedad en el curso de las experiencias del día a día, incluso para los médicos más perspicaces. Por otro lado, el pronóstico describe acontecimientos relativamente frecuentes. Por ejemplo, un gran porcentaje de pacientes con infarto agudo de miocardio mueren antes de abandonar el hospital.

Los factores pueden ser diferentes

Las variables asociadas con un mayor riesgo no son necesariamente las mismas que las que indican un peor pronóstico. A menudo, son muy diferentes para una enfermedad concreta. Por ejemplo, el número de factores de riesgo bien establecidos para la enfermedad cardiovascular (hipertensión, tabaquismo, dislipidemia, diabetes y antecedentes familiares de enfermedad coronaria) están inversamente relacionados con el riesgo de morir en el

booksmedicos.org

hospital después del primer infarto de miocardio (fig. 7-1) (1).

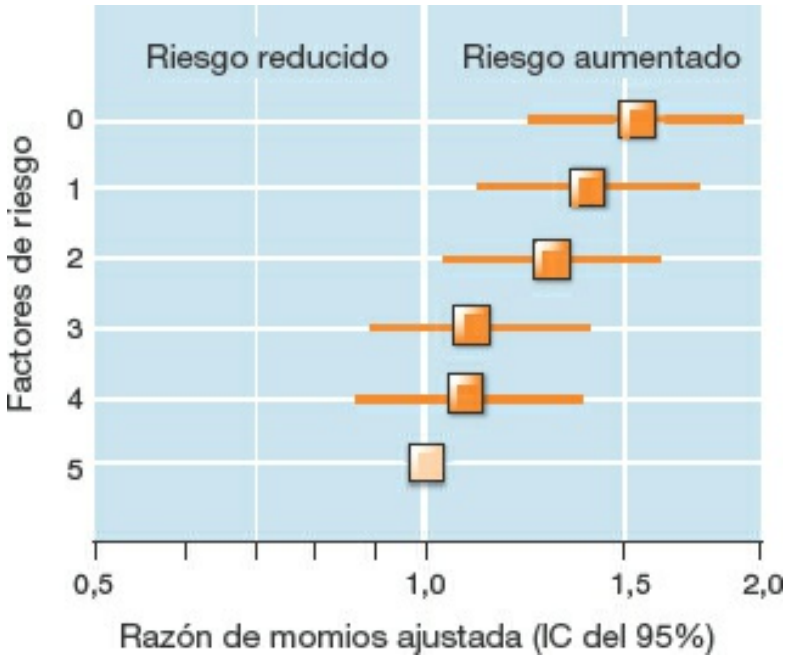


Figura 7-1 Factores pronósticos y de riesgo de infarto agudo de miocardio. Redibujado con autorización de Canto JC, Kiefe CI, Rogers WJ, et al. Number if coronary heart disease risk factors and mortality in patients with first myocardial infarction. JAMA 2011;306:2120-2127.)

A menudo, los médicos pueden realizar buenas estimaciones de pronósticos a corto plazo, a partir de su propia experiencia personal. Sin embargo, quizá les resulte más dificultoso discernir, sin el respaldo de la investigación, los diversos factores que se relacionan con el pronóstico a largo plazo o los complejos caminos en los que los factores pronóstico están relacionados entre sí.

CURSO CLÍNICO Y EVOLUCIÓN NATURAL DE LA ENFERMEDAD

El pronóstico puede describirse a partir del curso clínico o de la evolución natural de la enfermedad. El término **curso clínico** describe la evolución (pronóstico) de una enfermedad que ha seguido su curso bajo atención médica, y que se ha tratado mediante diversas formas que afectarán a la posterior evolución de los acontecimientos. Los pacientes suelen recibir asistencia médica en algún momento del curso evolutivo de su enfermedad, cuando presentan síntomas como dolor, alteraciones del crecimiento, desfiguración o conducta poco habitual. La diabetes mellitus de tipo 1, el

booksmedicos.org

carcinoma de pulmón y la rabia son ejemplos de ello. Una vez identificada la enfermedad, es probable que se trate.

El pronóstico de una enfermedad en la que no se produce intervención médica alguna se denomina **evolución natural** de la enfermedad, y describe la forma en que evolucionarán los pacientes si no se actúa sobre su enfermedad. Un gran número de afecciones no reciben asistencia médica, incluso en países con sistemas sanitarios avanzados. Estas permanecen sin ser identificadas porque son asintomáticas (p. ej., cánceres de próstata ocultos con un crecimiento lento) y, por lo tanto, no se reconocen durante la vida. En otros casos, como la osteoartritis, la depresión leve o la anemia de poca gravedad, los síntomas son una de las incomodidades ordinarias de la vida diaria y no una enfermedad y, en consecuencia, no buscan atención médica por estos motivos.

Ejemplo

El síndrome del colon irritable es una afección habitual que consiste en dolor abdominal y alteración de la defecación no causados por otras enfermedades. ¿Con qué frecuencia acuden al médico los pacientes con esta afección? En una cohorte inglesa de 3 875 personas que al inicio no tenían síndrome del colon irritable, el 15 % desarrolló el síndrome en los siguientes 10 años (2). De estos, sólo el 17 % consultó al médico de atención primaria por síntomas relacionados al menos una vez en 10 años, y el 4 % lo había consultado en el último año. En otro estudio, las características de los síntomas abdominales no constituyeron una razón para buscar o no asistencia médica para sus síntomas (3).

booksmedicos.org

Figura 7-2 ■ Diseño de un estudio de cohorte de riesgo.

ELEMENTOS DE LOS ESTUDIOS SOBRE PRONÓSTICO

La figura 7-2 muestra el diseño básico de un estudio de cohortes sobre pronóstico. En el mejor de los casos, los estudios sobre pronóstico se llevan a cabo en una población clínica o geográfica definida, inician la observación en un punto específico del tiempo a lo largo del curso de la enfermedad, incluyen un seguimiento de los pacientes durante un tiempo suficiente, y miden todos los resultados relevantes.

La muestra de pacientes

Al igual que los estudios sobre riesgo, los estudios sobre pronóstico se realizan de forma óptima con una población integrada por todas las personas con la enfermedad en cuestión en una región geográfica definida. De este modo, es posible asegurarse de que los pacientes descritos son una muestra sin sesgos de todos los pacientes de este tipo. En algunos países, la existencia de bases de datos con historias clínicas a nivel nacional permite la realización de estudios basados en la población.

Ejemplo

booksmedicos.org

Investigadores holandeses estudiaron el riesgo de complicaciones en el embarazo de mujeres con diabetes mellitus de tipo 1 (4). La muestra incluyó a las 323 mujeres holandesas con diabetes de tipo 1 que se habían quedado embarazadas a lo largo de 1 año y que habían sido atendidas en uno de los 118 hospitales del país. La mayoría de los embarazos se planificaron y, durante la gestación, la mayoría de las mujeres tomaron complementos de ácido fólico y mantuvieron un buen control glucémico. Sin embargo, las tasas de complicaciones en los recién nacidos fueron mucho más elevadas que en la población general. Se produjo morbilidad neonatal (una o más complicaciones) en el 80 % de los neonatos, y las tasas de malformaciones congénitas y los neonatos inusualmente grandes (macrosomía) fueron entre 3 y 12 veces superiores a las de la población general. Este estudio sugiere que sólo un óptimo control glucémico no fue suficiente para evitar las complicaciones durante la gestación en las mujeres con diabetes de tipo 1.

Incluso sin registros médicos nacionales es posible realizar estudios basados en la población. En EE.UU., la *Network of Organ Sharing* recoge datos de todos los pacientes con trasplantes, mientras que el programa de *Surveillance, Epidemiology, and End Results* (SEER) recoge los datos sobre la incidencia y supervivencia de todos los pacientes con cáncer de reciente aparición en varias de las mayores áreas del país, lo que comprende al 28 % de la población estadounidense. Para las preguntas de atención primaria, en EE.UU. y en otros sitios, las practicas individuales en las comunidades se han unido en las «redes de investigación de atención primaria» para recoger los datos de investigación sobre la atención de sus pacientes.

La mayoría de los estudios de pronóstico, sobre todo para las enfermedades poco habituales, se realizan en pacientes locales. En estos estudios es especialmente importante proporcionar información en la que puedan basarse los usuarios para decidir si los resultados pueden generalizarse a su propia situación: las características del paciente (p. ej., edad, gravedad de la enfermedad y comorbilidades), el sitio donde se encontraron (p. ej., práctica de atención primaria, hospitales comunitarios o centros de referencia) y cómo se obtuvo la muestra (p. ej., completa, aleatoria o muestreo por conveniencia). Con frecuencia, esta información es suficiente para establecer una generalización amplia, por ejemplo, en los estudios sobre la neumonía adquirida en la comunidad o la tromboflebitis en un hospital local.

Tiempo cero

Los estudios de cohortes sobre pronóstico deben partir de un punto común en el curso de la enfermedad, denominado **tiempo cero,** como el tiempo del

booksmedicos.org

inicio de los síntomas del diagnóstico o del inicio del tratamiento. Si se comienza la observación en diferentes momentos del curso de la enfermedad de los diversos pacientes de la cohorte, la descripción de su pronóstico carecerá de precisión y será difícil interpretar el desarrollo cronológico de la recuperación, las recidivas, la muerte y otros resultados, o bien inducirá a error. El término **cohorte de inicio** se utiliza para describir un grupo de pacientes que son reunidas al principio de su enfermedad.

Con frecuencia se describe individualmente el pronóstico del cáncer de acuerdo con el estado clínico del paciente (extensión de la diseminación) al inicio del seguimiento. Si se produce un cambio sistemático en la forma de establecer la etapa en el tiempo cero puede derivar en un pronóstico diferente para cada etapa, incluso si el curso de la enfermedad no se ha modificado para cada paciente de la cohorte. Se ha demostrado que esto ocurre durante la estadificación del cáncer, evaluación de la extensión de la enfermedad, con etapas más altas que corresponden a un cáncer más avanzado, para efectos de pronóstico y elección del tratamiento. La **fase de migración** se da cuando una nueva tecnología es capaz de detectar mejor la diseminación del cáncer que un método antiguo de estadificación. Los pacientes que solían clasificarse en una etapa inferior, con la tecnología más reciente, son clasificados como si estuvieran en una etapa mayor (más avanzada). La eliminación de los pacientes con una enfermedad más avanzada de las etapas más bajas aparentemente mejora el pronóstico para cada etapa, sin importar si el tratamiento fue más efectivo o el pronóstico global de estos pacientes mejoró. Los investigadores denominaron «fenómeno Will Rogers» a la migración de la etapa, en honor a un humorista estadounidense de principios de siglo, que, en referencia a las migraciones geográficas que se produjeron en EE.UU. durante la depresión económica de la década de 1930, afirmó: «Cuando los habitantes de Oklahoma se trasladaron a California, subió el promedio de inteligencia en ambos estados» (5).

Ejemplo

La tomografía por emisión de positrones (TEP), una prueba sensible para las metástasis, actualmente se utiliza para estadificar el cáncer no microcítico de pulmón. Los investigadores compararon las etapas del cáncer antes y después del uso generalizado de la TEP y observaron una disminución del 5,4 % en el número de pacientes con enfermedad en etapa III (cáncer diseminado dentro del tórax) y un aumento del 8,4 % en los pacientes con enfermedad en etapa IV (metástasis

booksmedicos.org

a distancia) (6). La estadificación con TEP se asoció con una mejor supervivencia en las etapas III y IV de la enfermedad, pero no en las etapas más tempranas. Los autores concluyeron que la migración de la etapa fue responsable de al menos una parte de la mejoría aparente en la supervivencia de los pacientes con cáncer de pulmón en etapas III y IV, que se presentó después de utilizar la estadificación con TEP.

Seguimiento

El seguimiento de los pacientes debe realizarse durante un período suficiente para que se produzca la mayoría de los resultados de interés clínicamente importantes; de lo contrario, la medida observada subestimará a la auténtica. La duración adecuada del seguimiento depende de la enfermedad. Para los estudios sobre la infección de herida quirúrgica, el período de seguimiento debe durar al menos unas cuantas semanas, y para los estudios sobre la aparición del SIDA y sus complicaciones en los pacientes con infección por VIH, el período debe durar varios años.

Resultados de la enfermedad

Las descripciones del pronóstico deben incluir todo el intervalo de manifestaciones de la enfermedad que se consideren importantes para los pacientes, lo que implica no sólo la muerte o el estado de enfermedad, sino también el dolor, el sufrimiento extremo y la incapacidad de cuidar de uno mismo o de llevar a cabo las actividades habituales. Muerte, enfermedad, malestar, incapacidad e insatisfacción son los cinco conceptos que resumen los resultados clínicos importantes (v. tabla 1-2).

En sus esfuerzos por ser «científicos», los médicos tienden a valorar los resultados precisos o medidos tecnológicamente por encima del resto, a veces, a expensas de la relevancia clínica. Tal como se ha comentado en el capítulo 1, los efectos clínicos que los pacientes no pueden percibir de forma directa, como la reducción radiológica del tamaño de un tumor, la normalización de los análisis bioquímicos de la sangre, la mejora de la fracción de eyección o los cambios en las pruebas serológicas, no constituyen fines clínicamente útiles de por sí. Es adecuado sustituir estos fenómenos biológicos por resultados clínicos sólo cuando se sepa que ambos están relacionados entre sí. Por lo tanto, en los pacientes con neumonía, la persistencia a corto plazo de las anomalías en la radiografía de tórax puede no ser alarmante si la fiebre del paciente ha cedido, ha recuperado su energía y su tos ha disminuido.

booksmedicos.org

Tabla 7-1

Una sencilla medida de la calidad de vida. La *Eastern Collaborative Oncology Group's Performance Scale*

Estado general	Definición
0	Asintomático
1	Sintomático, con plena capacidad de deambulación
2	Sintomático, en cama < 50 % del día
3	Sintomático, en cama > 50 % del día
4	Encamado
5	Moribundo

Adaptado con autorización de Oken MM, Creech RH, Tomey DC, et al. Toxicity and response criteria of the Eastern Oncology Group. Am J Clin Oncol 1982;5:649-655.

En la actualidad, en la investigación clínica se utilizan métodos para medir los resultados centrados en el paciente. La tabla 7-1 muestra un método simple para medir la calidad de vida utilizada en los estudios para el tratamiento del cáncer. También existen formas de medición para evaluar el estado funcional, la calidad de vida relacionada a la salud, el dolor y otros aspectos sobre el bienestar del paciente.

DESCRIPCIÓN DEL PRONÓSTICO

Es práctico resumir el curso de una enfermedad en forma de una sencilla medida: la proporción de personas que presentará un episodio en un período fijado. En la tabla 7-2 se muestran algunas de las medidas utilizadas para este fin. Tienen en común las mismas características básicas que la incidencia: episodios que surgen en una cohorte de pacientes a lo largo del tiempo.

Tabla 7-2

Medidas utilizadas habitualmente para describir el pronóstico

booksmedicos.org

Medida	Definición[a]
Supervivencia a los 5 años	Porcentaje de pacientes que sobrevive 5 años a partir de un punto determinado del curso de su enfermedad
Letalidad	Porcentaje de pacientes con una enfermedad que fallece a causa de esta
Mortalidad específica por una enfermedad	Número de personas por 10 000 (o 100 000) habitantes que muere a causa de una enfermedad específica
Respuesta	Porcentaje de pacientes que muestra indicios de mejora tras una intervención
Remisión	Porcentaje de pacientes que entra en una fase en que la enfermedad ya no es detectable
Recidiva	Porcentaje de pacientes que presenta una recidiva de la enfermedad tras un intervalo de tiempo sin ella

[a]El tiempo en observación se especifica o se supone que es lo bastante prolongado para que todos los acontecimientos que vayan a ocurrir se puedan observar.

Balance: sencillez frente a mayor información

Resumir el pronóstico con una sola medida presenta la virtud de la sencillez. Las medidas pueden conservarse en la memoria y comunicarse de forma sucinta. Como contrapartida, expresan una cantidad de información relativamente reducida. Medidas similares pueden ocultar grandes diferencias en el pronóstico.

La figura 7-3 muestra la supervivencia a los 5 años de pacientes con cuatro afecciones para cada una de las cuales un 10 % de los pacientes sigue con vida al cabo de ese período. Sin embargo, los cursos clínicos son, por lo demás, bastante diferentes en cuanto a aspectos que pueden ser muy

booksmedicos.org

importantes para los pacientes. La supervivencia inicial en los pacientes con aneurisma disecante es muy reducida, pero si sobreviven los primeros meses, el riesgo de morir se ve muy poco afectado por el hecho de haber sufrido el aneurisma (fig. 7-3 A). Los pacientes con cáncer no microcítico de pulmón localmente invasivo presentan una medida de mortalidad bastante constante a lo largo de los 5 años posteriores al diagnóstico (fig. 7-3 B). La vida de los pacientes con esclerosis lateral amiotrófica (ELA, la enfermedad de Lou Gehrig, una parálisis lentamente progresiva) con dificultad respiratoria no se ve amenazada de inmediato, pero a medida que la función neurológica continúa disminuyendo a lo largo de los años, la incapacidad para respirar sin asistencia causa la muerte (fig. 7-3 C). La figura 7-3 D se presenta como punto de referencia. Es únicamente a la edad de 100 años cuando las personas de la población general presentan una medida de supervivencia a los 5 años comparable a la de los pacientes con estas tres enfermedades.

Análisis de supervivencia

Cuando se interpretan los pronósticos, es preferible conocer la probabilidad media de que los pacientes con una afección concreta presenten un resultado en cualquier punto del tiempo. El pronóstico resumido en forma de medida no contiene esta información. Sin embargo, las cifras pueden resumir información sobre el tiempo medio hasta un acontecimiento para cualquier momento en el curso de la enfermedad. Por **acontecimiento** nos referimos a un resultado clínico dicotómico que puede ocurrir solamente una vez. En la exposición siguiente se adoptará la postura habitual de describir los resultados en términos de «supervivencia», si bien los mismos métodos se aplican al concepto inverso (tiempo hasta la muerte) y a cualquier otro resultado de interés, como recidiva del cáncer, curación de una infección, ausencia de síntomas o artritis inactiva.

booksmedicos.org

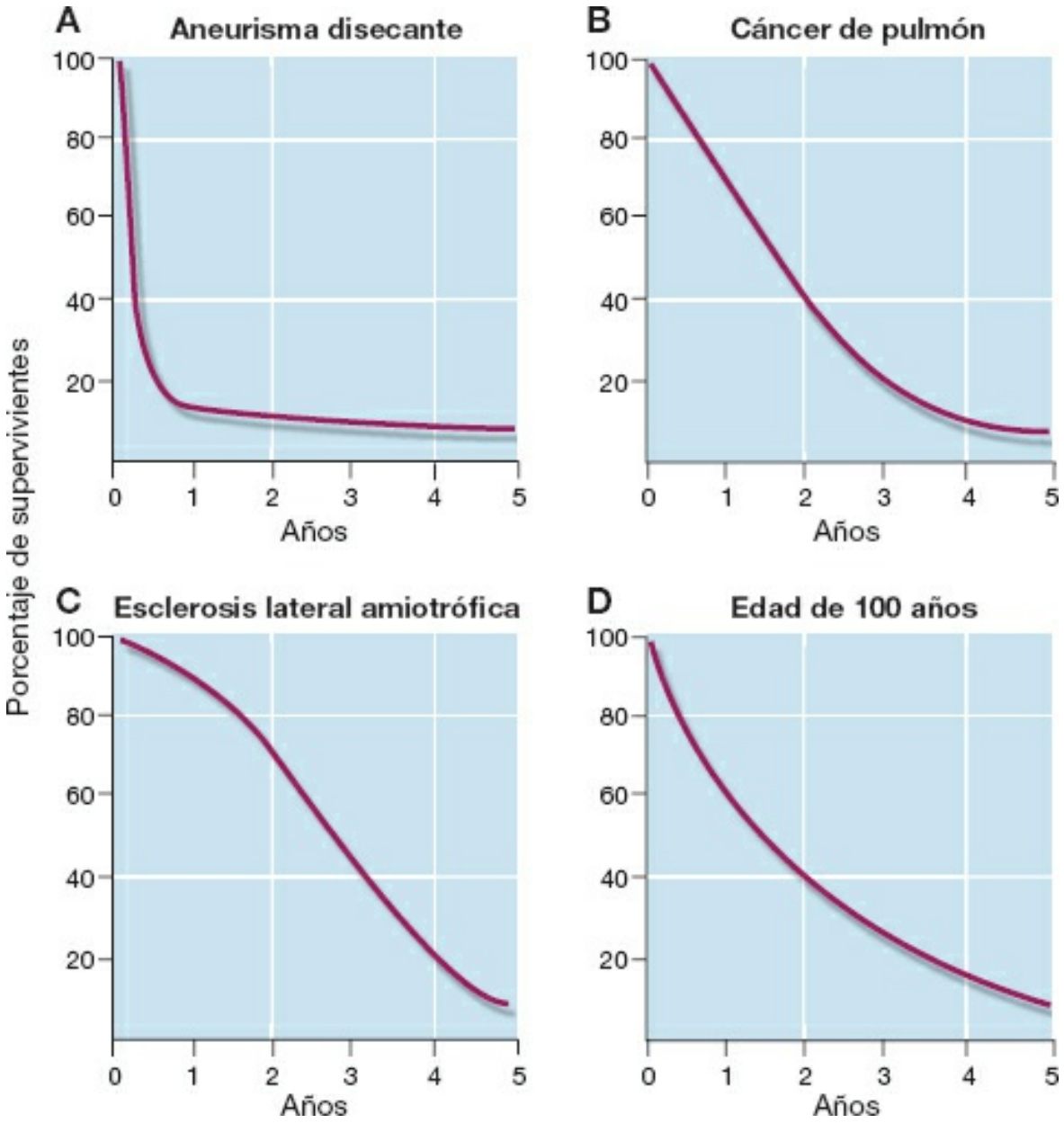


Figura 7-3 ■ Limitación de las tasas de supervivencia a los 5 años. Cuatro situaciones con la misma tasa de supervivencia a 5 cinco años: 10 %.

Supervivencia de una cohorte

La forma más directa de estudiar la supervivencia es constituir una cohorte de pacientes con la afección de interés que se encuentren en el mismo punto del curso de su enfermedad (como el inicio de los síntomas, el momento del diagnóstico o el inicio del tratamiento), y que se sometan a observación hasta que todos puedan presentar el resultado de interés o no. Para una cohorte pequeña, puede representarse el curso clínico de estos pacientes tal como se

booksmedicos.org

muestra en la figura 7-4 A. La supervivencia, en relación con el tiempo, se representa en forma de peldaños que corresponden a la muerte de cada uno de los 10 pacientes de la cohorte. Si se incrementara el número de pacientes, el tamaño de los peldaños disminuiría. Si se estudiara un gran número de pacientes, la figura se aproximaría a una curva lisa (fig. 7-4 B). Esta información podría utilizarse entonces para predecir el pronóstico de pacientes similares de un año a otro, o incluso de una semana a otra.

Por desgracia, obtener la información de esta forma no resulta práctico por varias razones. Sin duda, algunos pacientes abandonarían el estudio antes de que finalizara el período de seguimiento, quizá a causa de otras enfermedades, porque se mudaran a otro lugar fuera de la región del estudio o porque se sintieran insatisfechos con el mismo. Estos pacientes deberían excluirse de la cohorte, aunque se hubieran realizado grandes esfuerzos para obtener datos sobre ellos hasta el momento del abandono. Asimismo, sería necesario esperar a que todos los miembros de la cohorte hubieran alcanzado cada punto del seguimiento antes de poder calcular la probabilidad de supervivencia hasta ese punto. Ya que a menudo los pacientes están disponibles para un estudio durante un período de tiempo, en cualquier momento se produciría un seguimiento relativamente prolongado de los incluidos en el estudio en primer lugar, y se dispondría de una breve experiencia con los pacientes incluidos recientemente. El último paciente incluido en el estudio tendría que alcanzar cada año de seguimiento antes de disponer de cualquier información relacionada con la supervivencia hasta ese año.

Curvas de supervivencia

Para utilizar de forma eficaz todos los datos disponibles de cada paciente de la cohorte, se han creado los **análisis de supervivencia,** a fin de *calcular* la supervivencia de una cohorte a lo largo del tiempo. El análisis habitual se conoce como **método de Kaplan-Meier,** en honor a sus creadores. Los análisis de supervivencia pueden aplicarse a cualquier variable que sea dicotómica, y que se produzca una sola vez durante el seguimiento (p. ej., tiempo hasta un episodio coronario o hasta la recidiva de un cáncer). Un término más general, útil cuando se describe otro acontecimiento aparte de la supervivencia, es el **análisis de tiempo hasta el acontecimiento.**

booksmedicos.org

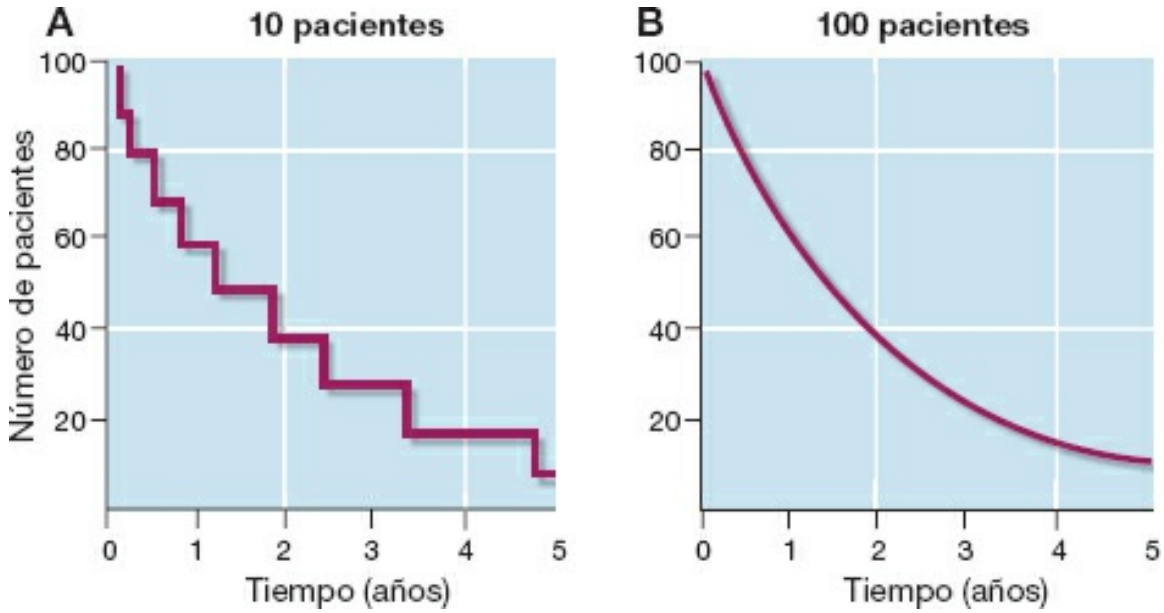


Figura 7-4 ■ Supervivencia de dos cohortes, una pequeña y otra grande, cuando todos los miembros son observados durante el período de seguimiento completo.

En la figura 7-5 se muestra una curva de supervivencia simplificada. En el eje de ordenadas se muestra la probabilidad de supervivencia estimada, y en el eje de abscisas el período transcurrido desde el inicio de la observación (tiempo cero).

La probabilidad de sobrevivir hasta cualquier punto del tiempo se calcula a partir de la probabilidad acumulada de sobrevivir en cada uno de los intervalos de tiempo que lo precedieron. Los intervalos de tiempo pueden reducirse tanto como sea necesario; en el método de Kaplan-Meier, los intervalos van de un acontecimiento a otro, por ejemplo, desde la muerte al acontecimiento anterior, independientemente de lo breve o prolongado que sea el intervalo resultante. La mayor parte del tiempo no fallece nadie, por lo que la probabilidad de sobrevivir es 1. Cuando un paciente muere, la probabilidad de sobrevivir en ese momento se calcula como el cociente entre el número de pacientes que sobrevive y el número de pacientes con riesgo de morir en ese momento. Los pacientes que ya han fallecido, han abandonado o no se ha realizado su seguimiento hasta aquel punto no presentan riesgo de morir, por lo que no se utilizan para calcular la supervivencia en aquel momento. La probabilidad de sobrevivir no cambia durante los intervalos en los que no fallece nadie, de forma que, en la práctica, la probabilidad de sobrevivir sólo se calcula de nuevo cuando se produce una muerte. Aunque la probabilidad asignada a cualquier intervalo determinado no es muy exacta,

booksmedicos.org

porque no ha sucedido nada o porque se ha producido sólo un acontecimiento en una cohorte de grandes dimensiones, la probabilidad global de sobrevivir hasta cada punto del tiempo (producto de todas las probabilidades previas) es notablemente exacta. Cuando en cualquier punto del tiempo se pierden pacientes para el estudio, pasan a calificarse como **pacientes censurados** y dejan de contabilizarse en el denominador a partir de ese momento.

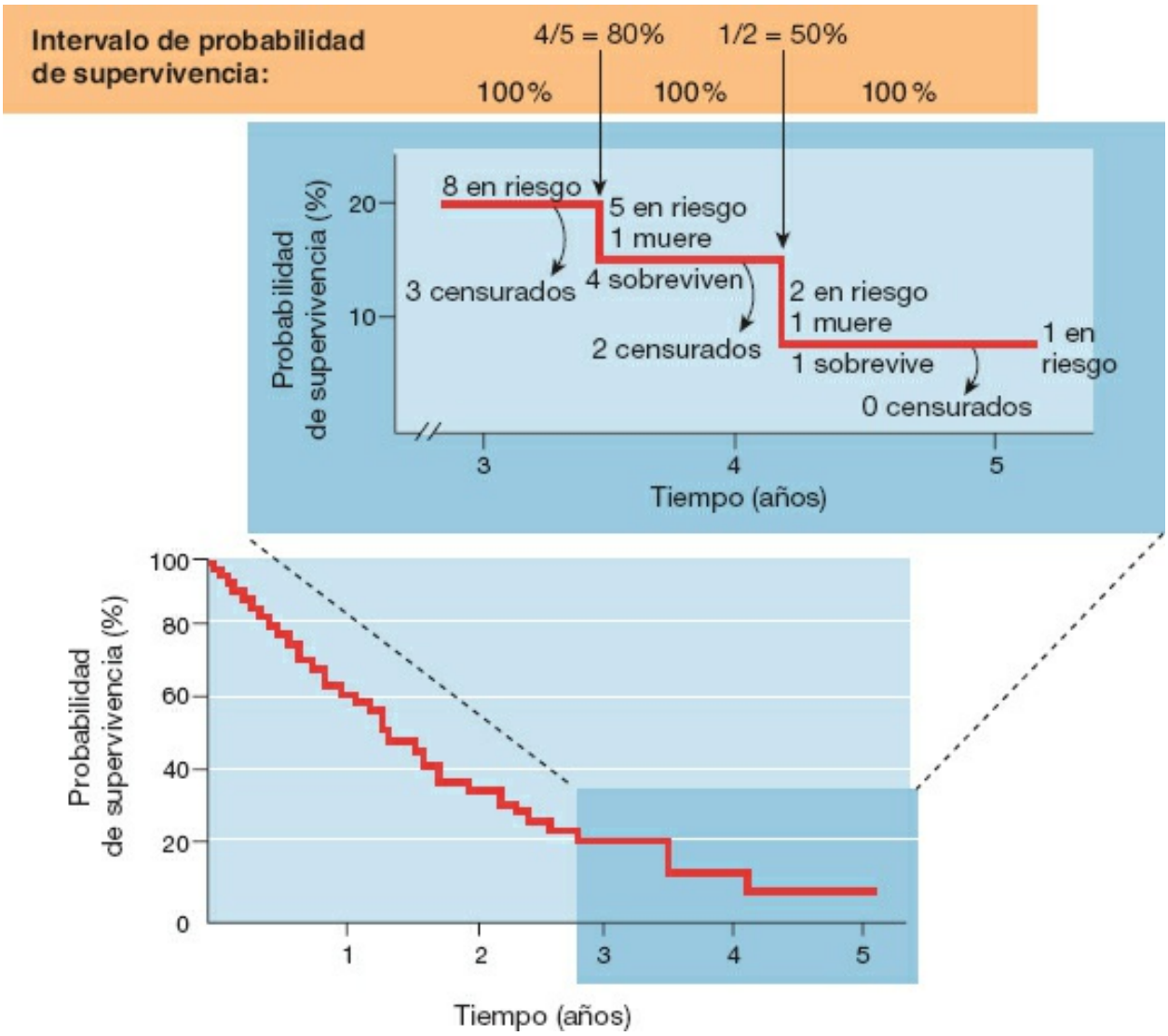


Figura 7-5 ■ Ejemplo de una curva de supervivencia. Una parte de esta se muestra de forma detallada.

Se presenta detalladamente una parte de la curva de supervivencia de la figura 7-5 (de los 3 a los 5 años desde el tiempo cero) para ilustrar los datos utilizados para estimar la supervivencia: pacientes en riesgo, pacientes que ya

booksmedicos.org

no están en riesgo (censurados) y pacientes que presentan resultados de interés en cualquier punto del tiempo.

Las variaciones en la curva de supervivencia básica incrementan la cantidad de información que estas contienen. La inclusión del número de pacientes en riesgo en varios puntos en el tiempo proporciona alguna idea sobre la contribución del azar a las tasas observadas, sobre todo hacia el final del seguimiento. En el eje vertical se puede mostrar la proporción de pacientes con el resultado de interés, en lugar de sin este; la curva resultante barrerá toda la superficie inicial hacia arriba, para desplazarse posteriormente hacia la derecha. La precisión de las estimaciones de supervivencia, que disminuye con el tiempo porque cada vez se observa un número menor de pacientes, puede demostrarse mediante intervalos de confianza en varios puntos del tiempo (v. cap. 11). En ocasiones se añaden marcas a las curvas de supervivencia para señalar cada vez que se censura a un paciente.

Interpretación de las curvas de supervivencia

Cuando se interpreten curvas de supervivencia, deben tenerse en cuenta varios puntos. En primer lugar, el eje vertical representa la probabilidad estimada de sobrevida para los miembros de la cohorte, no la incidencia acumulada de sobrevida si todos los miembros fueran seguidos.

En segundo lugar, los puntos de una curva de supervivencia son la mejor estimación, para un determinado conjunto de datos, de la probabilidad de supervivencia de los miembros de una cohorte. Sin embargo, la precisión de estas estimaciones depende, como todas las observaciones de muestras, del número de pacientes en los que se basa la estimación. Se puede estar más seguro de que las estimaciones situadas a la izquierda de la curva son firmes, porque durante ese tiempo existe un mayor número de pacientes en riesgo al inicio del seguimiento. Sin embargo, en la cola de la curva, situada a la derecha, puede que el número de pacientes sobre los que se basan los cálculos de supervivencia llegue a ser relativamente pequeño, debido a que los fallecimientos, los abandonos o las inclusiones tardías en el estudio den lugar a un número cada vez menor de pacientes sometidos a seguimiento para ese período de tiempo. Debido a ello, las estimaciones de supervivencia hacia el final del período de seguimiento son imprecisas, y pueden verse muy afectadas por lo que les sucede a un número bastante reducido de pacientes. Por ejemplo, en la figura 7-5, sólo un paciente permanecía en observación al cabo de 5 años. Si por casualidad dicho paciente muriera, la probabilidad de

booksmedicos.org

supervivencia caería del 8 % a 0. Claramente, esta lectura de los datos sería demasiado literal. Por consiguiente, las estimaciones de supervivencia en las colas de las curvas de supervivencia deben interpretarse con precaución.

Finalmente, el aspecto de algunas curvas de supervivencia da la impresión de que el resultado de interés se produce con más frecuencia al principio del seguimiento que más adelante, cuando la pendiente de la curva se aplana. Sin embargo, la impresión es falsa. A medida que pasa el tiempo, las tasas de supervivencia se aplican a un número cada vez menor de pacientes, lo que hace que la curva se aplane sin que varíe la tasa de resultados de interés.

Como en cualquier estimación, la estimación de Kaplan-Meier del tiempo transcurrido hasta el acontecimiento depende de suposiciones. Se asume que ser censurado no se relaciona con el pronóstico. En la medida de que esto no es cierto, el análisis de supervivencia puede producir estimaciones sesgadas en la supervivencia en las cohortes. El método de Kaplan-Meier puede no ser lo suficientemente exacto si hay riesgos que compiten entre sí, más de un tipo de evento resultante, y los resultados no son independientes el uno del otro, de modo que un acontecimiento cambia la probabilidad de experimentar otro. Por ejemplo, los pacientes con cáncer que desarrollan una infección relacionada con la quimioterapia agresiva y abandonan el estudio por esta razón pueden tener una probabilidad diferente de morir de cáncer. Existen otros métodos para estimar la incidencia acumulada ante la presencia de riesgos que compiten entre sí.

IDENTIFICACIÓN DE FACTORES PRONÓSTICOS

A menudo, se desea ir más allá de una sencilla descripción del pronóstico en un grupo de pacientes homogéneo para comparar los pronósticos en pacientes con características diferentes, es decir, con posibles factores pronósticos. En la misma figura se representan numerosas curvas de supervivencia, una para todos los pacientes con cada una de las características, lo que permite compararlas visual y estadísticamente.

Ejemplo

Los pacientes con carcinoma de células renales, al igual que muchos otros tipos de cáncer, tienen una probabilidad muy diferente de sobrevivir a lo largo de varios años después del diagnóstico. El

booksmedicos.org

pronóstico varía de acuerdo con las características del cáncer y el paciente, como la etapa (qué tan lejos se ha diseminado el cáncer, de estar limitado al riñón en un extremo hasta las metástasis a órganos distantes uno de otro), el grado (qué tan anómalas son las células cancerosas) y el estado funcional (qué tan bien pueden cuidar los pacientes de sí mismos). Un estudio combinó estas tres características en cinco grupos pronósticos (fig. 7-6) (7). En el grupo más favorable, más del 90 % de los pacientes estaban vivos a los 8 años, mientras que en el grupo menos favorable todos habían muerto a los 3 años. Esta información podría ser especialmente útil para ayudar a los pacientes y a los médicos a comprender lo que está por venir, y es mucho más informativa que simplemente decir a los pacientes que, «en general, el 70 % de los pacientes con carcinoma de células renales sobrevive 5 años».

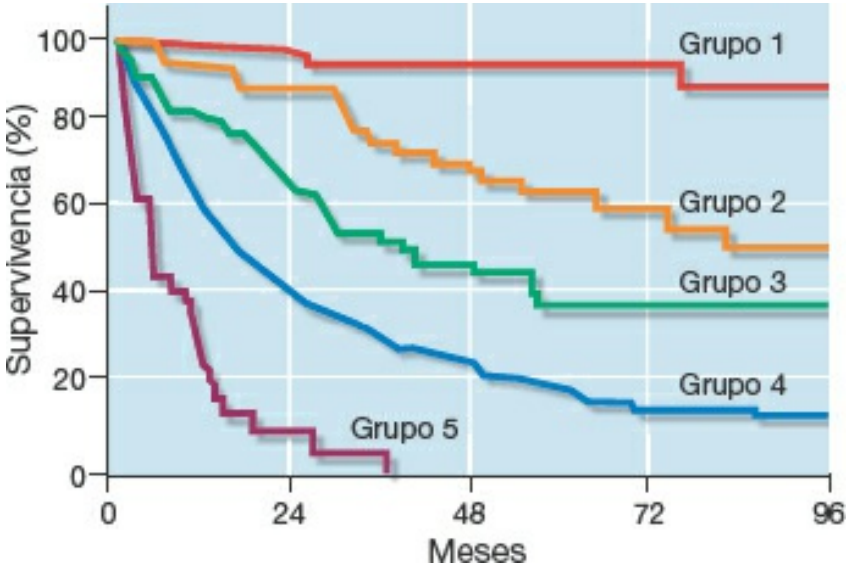


Figura 7-6 ■ Ejemplo de la estratificación pronóstica. Supervivencia de la cirugía en un paciente con cáncer de células renales de acuerdo con su estrato pronóstico. (Redibujado con autorización de Zisman A, Pantuck AJ, Dorey F, et al. Improved prognostification for renal cell carcinoma using an integrated staging system. J Clin Oncol 2001;19: 1649-1657.)

Los efectos relativos de las consecuencias de un factor pronóstico y otro pueden resumirse a partir de los datos en términos de un análisis del tiempo transcurrido hasta la presentación del evento por medio de una **razón de morbilidad** *(hazard ratio),* que es análoga a la razón de tasas. Asimismo, las curvas de supervivencia pueden compararse tras tener en cuenta otros factores relacionados con el pronóstico, de forma que se examine el efecto independiente de una sola variable.

SERIE DE CASOS

Una **serie de casos** es una descripción del curso de la enfermedad en un pequeño número de casos, a lo mucho unas pocas docenas. Un informe aún

booksmedicos.org

más pequeño, con menos de 10 pacientes, se denomina un **informe de caso.** Los casos generalmente se encuentran en una clínica o un centro de referencia y se siguen adelante en el tiempo para describir el curso de la enfermedad y hacia atrás para describir lo que se presentó antes.

Esos informes pueden realizar importantes contribuciones para entender la enfermedad, sobre todo mediante la descripción de experiencias iniciales en síndromes recién definidos o trastornos poco comprendidos. La razón para incluir una serie de casos en un capítulo sobre pronóstico es que pueden hacerse pasar por estudios de cohorte verdaderos a pesar que no tienen fortalezas comparables.

Ejemplo

Los médicos de urgencias atienden pacientes con mordeduras de serpientes de cascabel de América del Norte. Estas mordeduras son relativamente poco frecuentes, dificultando la realización de grandes estudios de cohortes sobre su curso clínico, por lo que los médicos deben confiar sobre todo en las series de casos. Un ejemplo es la descripción del curso clínico de los 24 niños tratados en un hospital infantil de California durante 10 años (8). A 19 niños se les inyectó el veneno, y fueron tratados agresivamente utilizando un antídoto. Tres fueron sometidos a tratamiento quirúrgico para retirar los restos de tejido blando o aliviar la presión de los tejidos. No hubo reacciones graves contra el antídoto y todos los pacientes dejaron el hospital sin deterioro funcional.

Los médicos que atienden a los niños con mordeduras de víboras de cascabel agradecerían esta y otras series de casos si no existiera otra información mejor para hacer su seguimiento, pero eso no quiere decir que la serie de casos proporcione una imagen completa y totalmente fiable de la atención de las mordeduras de serpiente. De todos los niños mordidos en esa región, algunos tuvieron una evolución tan favorable tras la mordedura que no fue necesario enviarlos al centro de referencia. Otros tenían una evolución tan desfavorable que se les trasladó rápidamente al hospital más cercano o incluso murieron antes de llegar. En otras palabras, las series de casos no describen el curso clínico de todos los niños desde el momento de la mordedura (el origen), sino de una muestra seleccionada de niños que casualmente fueron atendidos en ese hospital en particular. En efecto, una serie de casos describe el curso clínico de la prevalencia, y no necesariamente de una muestra representativa de los casos incidentales, por lo que son

booksmedicos.org

cohortes «falsas».

REGLAS DE PREDICCIÓN

Una combinación de variables aporta un pronóstico más preciso que cualquiera de dichas variables por separado. Como se comentó en el capítulo 4, las **reglas de predicción clínica** estiman la probabilidad de los resultados (ya sea de pronóstico o diagnóstico) en función de un conjunto de características de los pacientes definido por los antecedentes, la exploración física y las pruebas de laboratorio sencillas. Son «reglas» porque a menudo están vinculadas con recomendaciones sobre la evaluación diagnóstica o el tratamiento adicionales. Para hacer que las reglas de predicción sean factibles en el entorno clínico, estas dependen de los datos disponibles en la atención habitual de los pacientes y de la puntuación, la base para la predicción, que ha sido simplificada.

Ejemplo

La fibrilación auricular aumenta el riesgo de enfermedad cerebrovascular. Los coágulos se forman en la aurícula ante la ausencia de contracciones regulares y organizadas y, si se desplazan, pueden viajar hasta el cerebro causando enfermedad cerebrovascular por embolia. Esta complicación se puede prevenir con la anticoagulación pero con el riesgo de sangrado. Los médicos y los pacientes deben sopesar el riesgo de enfermedad cerebrovascular en ausencia de tratamiento contra el riesgo de complicaciones hemorrágicas por el tratamiento. Para facilitar esta decisión, se ha desarrollado y validado una regla de predicción clínica llamada CHADS2 (9). Se utiliza un conjunto de cinco observaciones clínicas fácilmente disponibles para separar a los pacientes con fibrilación auricular en seis estratos de riesgo con una diferencia en el riesgo de enfermedad cerebrovascular de 14 veces (tabla 7-3). A su vez, la agresividad de la anticoagulación está vinculada a la puntuación CHADS2 (10). Por ejemplo, las recomendaciones típicas basadas en el riesgo de enfermedad cerebrovascular y la eficacia de la anti-coagulación es que a los pacientes con una puntación CHADS2 de 0 no se les administra anticoagulación, aquellos con una puntuación de 1 se tratan con aspirina o anticoagulantes orales, y aquellos con una puntuación al menos de 2 son tratados con anticoagulación oral, el tratamiento más agresivo.

booksmedicos.org

Tabla 7-3

La puntuación CHADS2 y el riesgo de enfermedad cerebrovascular de acuerdo con esta puntuación

- ☐ Diagnóstico de insuficiencia cardíaca, anterior o actual (1 punto)
- ☐ Hipertensión con o sin tratamiento (1 punto)
- ☐ Edad ≥ 75 años (1 punto)
- ☐ Diabetes mellitus (1 punto)
- ☐ Prevención secundaria en pacientes con antecedente de enfermedad cerebrovascular isquémica, ataque isquémico transitorio o tromboembolia (2 puntos)

Puntuación CHADS2 = recuento total de puntos

Puntos CHADS2	Riesgo de enfermedad cerebrovascular por 100 personas-año
0	0,49
1	1,52
2	2,50
3	5,27
4	6,02
5-6	6,88

Reproducido como publicación en UpToDate, Calculation: Atrial Fibrillation CHADS (2) Score for Stroke Risk, with permission from MedCalc 3000 by Foundation Internet, Pittsburgh, PA.

Una regla de predicción clínica debe desarrollarse en un entorno y debe probarse en otros —con diferentes pacientes, médicos y prácticas habituales de atención—con el fin de garantizar que las predicciones son buenas para una gran cantidad de entornos y no sólo para lo que fue desarrollada, porque podría haber sido el resultado de las características particulares de ese entorno. Los datos utilizados para desarrollar la regla de predicción se denominan **conjunto de entrenamiento,** mientras que los datos utilizados para evaluar su validez se conocen como **conjunto de prueba,** que se utilizan para la **validación** de la regla de predicción.

El proceso de separar a los pacientes en grupos con diferente pronóstico, como en el ejemplo previo, se conoce como **estratificación pronóstica.** En este caso, la fibrilación auricular es la enfermedad y la enfermedad cerebrovascular el resultado. Este concepto es similar a la estratificación del

booksmedicos.org

riesgo (v. cap. 4), que divide a los pacientes en diferentes estratos de riesgo para desarrollar la enfermedad.

SESGO EN LOS ESTUDIOS DE COHORTES

En los estudios de cohorte de riesgo o pronóstico, los sesgos pueden alterar la descripción del curso de la enfermedad. Los sesgos también pueden crear diferencias aparentes entre los grupos cuando realmente no existen diferencias naturales o bien enmascarar las diferencias cuando estas existen. Estos sesgos también tienen sus homólogos en los estudios de casos y controles. Son una consideración separada de la confusión y la modificación del efecto (v. cap. 5).

Hay una variedad casi infinita de errores sistemáticos, a muchos de los cuales se les dan nombres específicos, pero algunos son más básicos. Pueden reconocerse más fácilmente cuando se sabe en qué momento es más probable que se presenten durante el curso de un estudio. Teniendo esto en mente, se describirán algunas posibilidades de sesgos en los estudios de cohorte y se comentarán en relación al siguiente estudio sobre pronóstico.

Ejemplo

La parálisis de Bell es la aparición repentina, unilateral e inexplicable de una debilidad en el rostro, en el área inervada por el nervio facial. Con frecuencia se desconoce la causa, aunque algunos casos se asocian con infecciones por herpes simple o herpes zóster, diabetes, embarazo y una serie de otros trastornos. ¿Cuál es el curso clínico? Investigadores holandeses revisaron a 1 701 pacientes con parálisis de Bell una vez al mes hasta que recuperaron la función o hasta 1 año después (11). La recuperación alcanzó su máximo en 3 semanas en el 85 % de los pacientes y de 3 a 5 meses en el 15 % restante. Para ese momento, el 71 % de los pacientes se había recuperado plenamente; para los demás, la pérdida restante de la función fue leve en el 12 %, moderada en el 13 % y grave en el 4 %. La recuperación se relacionó con la edad más joven, una menor parálisis al principio y el inicio temprano de la recuperación.

Cuando se piensa sobre la validez de este estudio, se debe considerar al menos lo siguiente.

Sesgo de muestreo

El **sesgo de muestreo** se produce cuando los pacientes de un estudio no son como los otros pacientes con la enfermedad. ¿Los pacientes de este estudio

booksmedicos.org

son como otros pacientes con parálisis de Bell? La respuesta depende de la perspectiva del usuario. Los pacientes eran del «área de Copenhague» y aparentemente estaban bajo el cuidado de un especialista en oídos, nariz y garganta, por lo que los resultados pueden generalizarse a otros pacientes derivados (mientras aceptemos que la parálisis de Bell en Dinamarca es un trastorno similar en otras partes del mundo). Sin embargo, los casos más leves podrían no haberse incluido porque los trataron médicos locales y se recuperaron con rapidez o no recibieron atención médica, lo que limita su aplicabilidad en el ámbito de la atención primaria.

El sesgo de muestreo también puede ser engañoso cuando se compara el pronóstico entre los grupos y el muestreo ha producido grupos que son sistemáticamente diferentes con respecto al pronóstico, incluso antes de considerar el factor de interés. En el ejemplo de la parálisis de Bell, los pacientes mayores podrían haber tenido un pronóstico peor porque eran los que tenían una infección subyacente por el virus del herpes y no por su edad.

¿Esto no es un factor de confusión? Hablando estrictamente, no lo es, porque el propósito del estudio es la predicción y no identificar las «causas» independientes de recuperación. Además, no es posible considerar fenómenos como la gravedad de la parálisis o el inicio de la recuperación como causas porque probablemente son parte de la cadena de acontecimientos que llevan de la enfermedad a la recuperación. Pero en la medida que se quiere demostrar que los factores pronósticos son predictores independientes del resultado, se pueden utilizar las mismas estrategias que se usan para la confusión (v. cap. 5) para establecer la independencia.

Sesgo de migración

El **sesgo de migración** está presente cuando algunos pacientes abandonan el estudio durante el seguimiento y son sistemáticamente diferentes a los que permanecen en el estudio. Es un tema frecuente que algunos de los miembros de la cohorte original abandonen el estudio después de un tiempo. (A los pacientes se les garantiza este derecho como parte de una conducta ética en la investigación con seres humanos).

No se produciría ningún sesgo si estos **abandonos** se produjeran de forma aleatoria, de modo que las características de los antiguos miembros de un grupo fueran, de promedio, similares a las de los antiguos miembros del otro grupo. Este hecho no cambiaría sin importar si el número de abandonos es elevado o si es similar en ambos grupos, pero a menudo las características de

booksmedicos.org

los pacientes perdidos no son las mismas que los que permanecen en el estudio. El abandono tiende a estar relacionado con el pronóstico. Por ejemplo, los pacientes con parálisis de Bell con una evolución especialmente buena o mala pueden ser más propensos a abandonar el estudio, al igual que aquellos que necesitan atención por otras enfermedades, para quienes las consultas adicionales relacionadas con el estudio serían molestas. Esto podría distorsionar los resultados principales (descriptivos) del estudio, la tasa y la totalidad de la recuperación. Si el estudio también pretende identificar los factores pronósticos (p. ej., la recuperación de los pacientes mayores en comparación con los más jóvenes), también puede estar sesgado por el abandono de los pacientes por las mismas razones.

El sesgo de migración puede ser considerado como un ejemplo del sesgo de selección debido a que los pacientes que todavía estaban en el estudio cuando los desenlaces fueron medidos, se seleccionaron de todos aquellos que iniciaron el estudio. Este sesgo puede considerarse un ejemplo del sesgo de medición, ya que los pacientes que migran fuera del estudio ya no están disponibles cuando se miden los resultados.

Sesgo de medición

El **sesgo de medición** está presente cuando no se evalúa el resultado de manera similar en los miembros de la cohorte. En el estudio de la parálisis de Bell todos los miembros de la cohorte fueron examinados mediante un protocolo común cada mes hasta que dejaron de mejorar, descartando esta posibilidad. Si se hubiera permitido a los propios pacientes y a los médicos decidir sobre la exploración, cuándo y cómo realizarla, habría disminuido la confianza en la descripción del tiempo para la recuperación en su totalidad. El sesgo de medición también entra en juego si se comparan los grupos pronósticos y se tiene una probabilidad sistemáticamente mejor de detectar los resultados de los pacientes de un grupo y no de los del otro grupo. Algunos resultados, como la muerte, las complicaciones cardiovasculares y los cánceres más adversos, son tan obvios que es improbable que se pasen por alto. Sin embargo, con resultados menos claros, como la causa específica de la muerte, una enfermedad subclínica, los efectos secundarios o una incapacidad, puede producirse un sesgo de medición debido a diferencias en los métodos con los que se busca o se clasifica el resultado.

El sesgo de medición puede reducirse de tres maneras: *a)* examinar igual a todos los miembros de la cohorte para determinar los acontecimientos de

booksmedicos.org

resultado; *b)* si se hacen comparaciones sobre los grupos de pronóstico, garantizar que los investigadores desconocen el grupo al que pertenece cada paciente, y *c)* establecer reglas cuidadosas para decidir si se ha producido un acontecimiento de resultado (y seguir las reglas). Para ayudar a los lectores a comprender la magnitud de este tipo de sesgos en cierto estudio, es una práctica habitual incluir, con los informes del estudio, un diagrama de flujo que describa cómo y por qué cambió el número de participantes conforme avanzó el estudio. También es útil comparar las características de los pacientes dentro y fuera del estudio después del muestreo y el seguimiento.

Sesgos por una clasificación errónea no diferencial

Hasta ahora, se ha discutido sobre cómo los resultados del estudio pueden estar sesgados cuando hay diferencias sistemáticas en la clasificación de los grupos de exposición o enfermedad. Pero también pueden presentarse sesgos si hay una clasificación errónea no diferencial, es decir, se produce de manera similar en los grupos que serán comparados. En este caso, el sesgo se inclina hacia no encontrar un efecto.

Ejemplo

Cuando se evalúa el consumo de cigarrillos simplemente preguntando a la gente si fuma, hay una clasificación errónea importante en relación con el patrón de referencia, como la presencia o ausencia de productos de tabaco en la saliva. Sin embargo, en un estudio de cohorte sobre el consumo de cigarrillos y la enfermedad cardíaca coronaria (ECC), la clasificación errónea del tabaquismo no podía ser diferente en las personas que desarrollaron o no ECC, porque no se conocía el resultado en el momento que se evaluó la exposición. Aun así, en la medida que se clasifica erróneamente el tabaquismo, se reduce cualquier diferencia que podría haber existido en la tasa de ECC entre los fumadores y no fumadores si todos los pacientes se hubieran clasificado de forma correcta, aumentando la probabilidad de un efecto «nulo». En el extremo, si se clasificara el consumo de tabaco totalmente al azar, podría no existir asociación alguna entre el tabaquismo y la ECC.

¿ES REALMENTE IMPORTANTE EL SESGO?

La epidemiología clínica no es un juego que consiste en ver quién encuentra los errores. Por el contrario, se ha concebido para caracterizar la credibilidad de un estudio, de forma que los médicos puedan decidir cuánto confían en sus resultados a la hora de apostar fuerte por una decisión sobre sus pacientes.

booksmedicos.org

Sería irresponsable ignorar los resultados de los estudios que cumplen estrictas normas de calidad, de la misma forma que las decisiones clínicas no deben estar sujetas a los resultados de los estudios que son poco sólidos.

Teniendo esto en cuenta, no basta con admitir que un estudio podría estar sesgado. Es necesario determinar si el sesgo está o no realmente presente en el estudio que nos atañe. Más allá de esta consideración, hay que decidir si las consecuencias del sesgo son lo bastante graves como para cambiar las conclusiones de una forma clínicamente importante. Si no perjudica demasiado a las conclusiones del estudio, la presencia del sesgo tendrá pequeñas repercusiones prácticas, y el estudio seguirá siendo de utilidad.

ANáLISIS DE SENSIBILIDAD

Una forma de decidir si el sesgo podría cambiar las conclusiones de un estudio es hacer un **análisis de sensibilidad,** es decir, demostrar qué tan grandes o pequeños podrían haber sido los resultados, observados bajo diferentes hipótesis sobre los datos perdidos o mediciones potencialmente sesgadas. Un **análisis del mejor/peor caso** pone a prueba los efectos de las hipótesis más extremas posibles, pero en la mayoría de las situaciones es una prueba extraordinariamente severa para los efectos del sesgo. Con mayor frecuencia, el análisis de sensibilidad prueba los efectos de valores algo improbables, como en el siguiente ejemplo.

Ejemplo

La poliomielitis se ha erradicado en varias partes del mundo, pero los efectos tardíos de la infección aún persisten. Algunos pacientes desarrollan síndrome de la pospoliomielitis (debilidad muscular, dolor y fatiga) muchos años después de la infección original. Para describir la tasa de síndrome pospoliomielitis, algunos investigadores identificaron una cohorte de 939 pacientes que tuvieron poliomielitis en la década de 1950 (12). Casi 40 años después fueron capaces de obtener información sobre los síntomas de 551 miembros de la cohorte, y en ellos, la tasa de síndrome pospoliomielitis fue de 137 en 551 (25 %). Es comprensible que después de tantos años 388 pacientes no pudieron ser objeto de seguimiento.

Si los miembros perdidos de la cohorte tenían tasas diferentes de síndrome pospoliomielitis de los que fueron objeto del seguimiento, ¿cuánto podrían haber sesgado la tasa observada de síndrome pospoliomielitis? Si los pacientes que no fueron objeto de seguimiento tuvieran el doble de probabilidades de adquirir el síndrome pospoliomielitis en comparación con los demás, la tasa verdadera hubiera sido (137 + 194)/939 = 35 %. (Es decir, los 137 que se sabe que tienen el síndrome más el 50 % de aquellos que no fueron objeto de seguimiento dividido entre todos los miembros de la cohorte original). Si los pacientes que no fueron objeto de seguimiento tuvieran la

booksmedicos.org

mitad de probabilidades de adquirir el síndrome, la tasa verdadera hubiera sido (137 + 48)/939 = 20 %. Por lo tanto, incluso con una improbable diferencia en las tasas de síndrome pospoliomielitis en los pacientes ausentes, la tasa verdadera todavía habría estado en el intervalo del 20-35 %, una aproximación útil para el uso clínico.

Se podrían haber asumido más o menos diferencias extremas en los pacientes ausentes para explorar qué tan «sensible» sería el resultado del estudio a los miembros ausentes de la cohorte. El análisis de sensibilidad es un método útil para estimar cuánto podrían estar afectados los resultados por los diferentes tipos de sesgos en estudios de todo tipo, estudios de cohorte y de casos y controles de riesgo o pronóstico, la exactitud o validez de las pruebas diagnósticas o los ensayos clínicos sobre la efectividad del tratamiento o la prevención.

Preguntas de revisión

Lea los siguientes enunciados y señale la mejor opción.

7-1. En un estudio sobre el riesgo de cáncer esofágico en pacientes con esófago de Barrett (una lesión premaligna), ¿cuál de los siguientes momentos en el curso de la enfermedad es el mejor ejemplo del tiempo cero?

A. El diagnóstico de esófago de Barrett para cada paciente.
B. La muerte de cada paciente.
C. El diagnóstico de cáncer de esófago para cada paciente.
D. La fecha cuando el primer paciente entró en el estudio.
E. La fecha cuando ningún paciente permaneció en el estudio.

7-2. Un estudio de cohorte describe la recurrencia de convulsiones en 1 año en niños hospitalizados por una primera convulsión febril. Se comparó la recurrencia en los niños que tuvieron una infección en comparación con la vacunación como la causa subyacente de la fiebre en el momento de la primera convulsión. Un año después algunos de los niños ya no estaban en el estudio cuando se evaluó el resultado (una segunda convulsión). ¿Cuál de los siguientes podría haber tenido el mayor efecto sobre los

booksmedicos.org

resultados del estudio?

A. La razón por la que los niños abandonaron el estudio.
B. El momento en el curso de la enfermedad en que los niños abandonaron el estudio.
C. Si el abandono se relacionó con el pronóstico.
D. Si el número de niños que abandonó el estudio fue similar entre los grupos.

7-3. En un estudio de cohorte sobre la atención de cáncer de próstata se comparan las tasas de incontinencia en pacientes que fueron tratados con cirugía en comparación con la atención médica. La incontinencia fue evaluada a partir de la revisión de historias clínicas. ¿Cuál de los siguientes *no* es un ejemplo del sesgo de medición?

A. Era más probable que los hombres le informaran a su cirujano acerca de la incontinencia.
B. Era menos probable que los cirujanos consignaran las complicaciones de su cirugía en el registro.
C. Era más probable que los hombres sometidos a cirugía tuvieran visitas de seguimiento.
D. Los revisores de los registros utilizaron su criterio para decidir si la incontinencia estaba presente o no.
E. Las tasas de incontinencia fueron mayores en los hombres sometidos a cirugía.

7-4. Se ha creado una regla de predicción clínica para clasificar el pronóstico de la neumonía adquirida en la comunidad. ¿Cuál de las siguientes caracteriza a una regla de predicción clínica como esta?

A. Es sencillo calcular la puntuación.
B. Los datos clínicos se encuentran fácilmente disponibles.
C. Se incluyen múltiples factores pronósticos.
D. Los resultados se usan para guiar aún más la atención del paciente.
E. Todas las anteriores.

7-5. Un estudio describe el curso clínico de pacientes que sufren una enfermedad neurológica poco habitual. Los pacientes se identifican en

booksmedicos.org

un centro de referencia especializado en esta enfermedad. Se revisan sus historias médicas en busca de las características de los pacientes y sus tratamientos. ¿Cuál de los siguientes describe mejor este tipo de estudio?

A. Estudio de cohorte.
B. Estudio de casos y controles.
C. Serie de casos.
D. Estudio transversal.
E. Ensayo controlado y aleatorizado.

7-6. Un estudio utilizó un análisis del tiempo transcurrido hasta el acontecimiento para describir la supervivencia desde el diagnóstico en 100 pacientes con insuficiencia cardíaca congestiva. Para el tercer año, 60 pacientes habían sido censurados. ¿Cuál de las siguientes no podría haber sido una razón para que uno de estos pacientes hubiera sido censurado?

A. El paciente murió por otra causa antes del tercer año.
B. El paciente decidió no continuar en el estudio.
C. El paciente desarrolló otra enfermedad que podría haber sido mortal.
D. El paciente había sido reclutado en el estudio menos de 3 años.

7-7. ¿Cuál de los siguientes tipos de estudio no podría utilizarse para identificar los factores pronósticos?

A. Estudio de prevalencia.
B. Análisis del tiempo transcurrido hasta el acontecimiento.
C. Estudio de casos y controles.
D. Estudio de cohortes.

7-8. ¿Cuál de los siguientes describe mejor la información de una curva de supervivencia?

A. Una estimación no sesgada de la supervivencia incluso si algunos pacientes abandonan el estudio.
B. La probabilidad estimada de supervivencia desde el tiempo cero.
C. La proporción de la cohorte que continúa viva al final del seguimiento.

booksmedicos.org

D. La tasa de abandono del estudio de los miembros originales de la cohorte.

E. La supervivencia acumulativa de una cohorte a lo largo del tiempo.

7-9. ¿Cuál de los siguientes es la muestra más apropiada para un estudio sobre pronóstico?

A. Miembros de la población general.

B. Pacientes tratados en atención primaria en la comunidad.

C. Pacientes ingresados en un hospital comunitario.

D. Pacientes derivados a un especialista.

E. Depende de quién utilizará los resultados del estudio.

7-10. Los investigadores desean describir el curso clínico de la esclerosis múltiple. Para ello utilizan un ensayo clínico ya completado en el que los pacientes utilizados como control recibieron la asistencia habitual. Los pacientes del ensayo se habían identificado en centros de referencia, se habían incluido en el momento del diagnóstico y habían cumplido rigurosos criterios de inclusión. Al cabo de 10 años todos los pacientes permanecían en observación, y un 40 % de ellos todavía era capaz de andar. ¿Cuál de los siguientes limita más la credibilidad del estudio?

A. Tiempo cero desigual.

B. Capacidad de generalización.

C. Sesgo de medición.

D. Sesgo de migración.

E. Fallo en la utilización de los métodos de análisis del tiempo transcurrido hasta el acontecimiento.

7-11. Se han creado muchas reglas de predicción clínica para evaluar la gravedad de la neumonía adquirida en la comunidad. ¿Cuál de las siguientes es la razón más importante para elegir una de ellas para el uso rutinario?

A. La regla de predicción clasifica a los pacientes en grupos con pronósticos muy diferentes.

B. La regla de predicción ha sido validada en diferentes entornos.

C. Se han incluido muchas variables en la regla.

booksmedicos.org

D. Los factores pronósticos incluyen las pruebas diagnósticas de la ciencia actual.

E. La puntuación se calcula utilizando ordenadores.

7-12. En un estudio sobre los pacientes que fumaron y desarrollaron enfermedad arterial periférica, la *hazard ratio* de amputación en los pacientes que siguieron fumando, en comparación con los que dejaron de fumar, es de 5. ¿Cuál de los siguientes caracteriza mejor a esta medida de la fuerza de la morbilidad?

A. Para este estudio, es la tasa de continuar fumando dividida entre la tasa de dejar de fumar.

B. Puede estimarse a partir de un estudio de casos y controles sobre el tabaquismo y la amputación.

C. No puede ser ajustado por la presencia o ausencia de otros factores relacionados con el pronóstico.

D. Proporciona información similar al riesgo relativo.

E. Se calcula a partir de la incidencia acumulada de amputación en fumadores y no fumadores.

7-13. En el análisis del tiempo transcurrido hasta el acontecimiento, el evento:

A. Puede ocurrir solamente una vez.

B. Es dicotómico.

C. Tanto A como B.

D. Ni A ni B.

Encontrará las respuestas en el apéndice A.

BIBLIOGRAFÍA

1. Canto JG, Kiefe CI, Rogers WJ, et al. Number if coronary heart disease risk factors and mortality in patients with first myocardial infarction. JAMA 2011;306:2120–2127.
2. Talley NJ, Zinsmeister AR, Van Dyke C, et al. Epidemiology of colonic symptoms and the irritable bowel syndrome. Gastroenterology 1991;101:927–934.
3. Ford AC, Forman D, Bailey AG, et al. Irritable bowel syndrome: A 10-year natural history of symptoms and factors that influence consultation behavior. Am J Gastroenterol 2007; 103:1229–1239.
4. Evers IM, de Valk HW, Visser GHA. Risk of complications of pregnancy in women with type 1

booksmedicos.org

diabetes: Nationwide prospective study in the Netherlands. BMJ 2004;328:915– 918.
5. Feinstein AR, Sosin DM, Wells CK. The Will Rogers phenomenon: stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. N Engl J Med 1985;312:1604–1608.
6. Chee KG, Nguyen DV, Brown M, et al. Positron emission tomography and improved survival in patients with lung cancer. The Will Rogers phenomenon revisited. Arch Intern Med 2008;168:1541–1549.
7. Zisman A, Pantuck AJ, Dorey F, et al. Improved prognostification for renal cell carcinoma using an integrated staging system. J Clin Oncol 2001;19:1649–1657.
8. Shaw BA, Hosalkar HS. Rattlesnake bites in children: antivenin treatment and surgical indications. J Bone Joint Surg Am 2002;84-A(9):1624.
9. Gage BF, Waterman AD, Shannon W, et al. Validation of clinical classification schemes for predicting stroke. Results from the National Registry of Atrial Fibrillation. JAMA 2001; 285:2864–2870.
10. Go AS, Hylek EM, Chang Y, et al. Anticoagulation therapy for stroke prevention in atrial fibrillation. How well do randomized trials translate into clinical practice? JAMA 2003;290:2685–2692.
11. Peitersen E. Bell's palsy: the spontaneous course of 2,500 peripheral facial nerve palsies of different etiologies. Acta Otolaryngol 2002;(549):4–30.
12. Ramlow J, Alexander M, LaPorte R, et al. Epidemiology of post-polio syndrome. Am J Epidemiol 1992;136:769–785.

booksmedicos.org

Capítulo 8

Diagnóstico

Para la mente, las apariencias son de cuatro tipos. Las cosas son lo que aparentan ser; o no son ni aparentan ser; o son y no lo aparentan; o no son aunque aparentan serlo. El cometido del hombre sabio es distinguir entre todos estos casos.

—Epicteto [†]
siglo II d.C.

PALABRAS CLAVE

Prueba diagnóstica
Verdadero positivo
Verdadero negativo
Falso positivo
Falso negativo
Estándar de oro
Patrón de referencia
Patrón de comparación
Sensibilidad
Especificidad
Punto de corte
Curva ROC (curva de eficacia diagnóstica)
Espectro
Sesgo
Valor predictivo
Valor predictivo positivo
Valor predictivo negativo
Probabilidad posterior (posprueba)

booksmedicos.org

Exactitud
Prevalencia
Probabilidad previa (preprueba)
Razón de verosimilitud
Probabilidad
Momios o posibilidades *(odds)*
Momios preprueba
Momios posprueba
Pruebas en paralelo
Pruebas en serie
Reglas de predicción clínica
Reglas de toma de decisiones diagnósticas

Los médicos dedican una gran parte del tiempo a diagnosticar los síntomas o las anomalías de sus pacientes. En general, suelen llegar a un diagnóstico después de llevar a cabo diversas pruebas diagnósticas. En el momento de interpretar los resultados de dichas pruebas, los médicos deben estar familiarizados con algunos principios básicos, que se tratarán en este capítulo.

Habitualmente, se entiende por **prueba diagnóstica** aquella realizada en un laboratorio, si bien los principios que se tratan en este capítulo pueden aplicarse igualmente a la información clínica obtenida mediante el interrogatorio, la exploración física y las pruebas de diagnóstico por imagen. Asimismo, también son válidos cuando, como prueba diagnóstica, se utiliza una constelación de hallazgos. Así pues, puede hablarse del valor de los síntomas neurológicos prodrómicos, la cefalea, las náuseas y los vómitos, en el diagnóstico de la migraña clásica, o de la hemoptisis y la pérdida de peso en un fumador de cigarrillos, como indicadores de cáncer de pulmón.

SIMPLIFICACIÓN DE LOS DATOS

En el capítulo 3 señalamos que las mediciones clínicas, entre ellas los datos de las pruebas diagnósticas, se expresan en escalas nominales, ordinales o de intervalo. Independientemente del tipo de datos que proporcionen las pruebas diagnósticas, los médicos suelen simplificarlos para que sean útiles en la práctica. La mayor parte de las escalas ordinales son ejemplos de este proceso de simplificación. Los soplos cardíacos pueden oscilar entre muy fuertes e inaudibles, pero para la toma de decisiones clínicas no es necesario expresar las gradaciones sutiles en la intensidad de los soplos; con una sencilla escala ordinal (grados I-VI) es suficiente. Con mucha más frecuencia, los datos

booksmedicos.org

complejos se reducen a una simple dicotomía (p. ej., presente/ausente, normal/anormal o sano/enfermo). Concretamente, esto se hace cuando los resultados de las pruebas se utilizan para contribuir a determinar las decisiones sobre el tratamiento. Con cualquier resultado de una prueba, las decisiones terapéuticas se limitan a una u otra elección: o se inicia el tratamiento o este queda diferido. Cuando existen categorías de tratamiento de acuerdo con el resultado de la prueba, los datos se tratan de una manera ordinal.

---

†Un médico del s. XIX y defensor del «método numérico» (basado en las cuentas, no en las impresiones) para comprender la historia natural de las enfermedades como la fiebre tifoidea.

Ejemplo

La forma en que se utilizan los datos de la presión arterial para tomar decisiones acerca del tratamiento es un ejemplo de cómo puede simplificarse la información por razones clínicas prácticas. La tensión arterial suele acotarse hasta el milímetro de mercurio más cercano, es decir, en una escala de intervalo. Sin embargo, la mayoría de las directrices terapéuticas para la hipertensión, como las del Joint National Committee on the Detection, Evaluation, and Treatment of Hypertension (1), eligen un nivel concreto (p. ej., 140 mm Hg de tensión sistólica o 90 mm Hg de tensión diastólica), a partir del cual recomiendan iniciar un tratamiento farmacológico. Con ello, transforman los datos de intervalo en datos dicotómicos. Para seguir con el ejemplo, que a continuación convierte los datos a una escala ordinal, el Joint National Committee también recomienda que los médicos elijan un plan terapéutico dependiendo de que la tensión arterial del paciente corresponda a «prehipertensión» (tensión sistólica de 120-139 mm Hg o diastólica de 80-89 mm Hg), «hipertensión de fase 1» (sistólica de 140-159 mm Hg o diastólica de 90-99 mm Hg) o «hipertensión de fase 2» (sistólica ≥ 160 mm Hg o diastólica ≥ 100 mm Hg).

EXACTITUD DEL RESULTADO DE UNA PRUEBA

Establecer diagnósticos es un proceso imperfecto que da como resultado una probabilidad, en lugar de la seguridad de haber acertado. Antes de obtener un diagnóstico clínico, el grado de certeza o de incertidumbre del médico acerca de dicho diagnóstico se expresa con términos como «descartar» o «posible». Cada vez más, los médicos expresan la posibilidad de que un paciente sufra una enfermedad en forma de probabilidad. En este caso, corresponde al médico estar familiarizado con las relaciones matemáticas que existen entre las propiedades de las pruebas diagnósticas y con la información que estas proporcionan en varias situaciones clínicas. En muchos casos, la comprensión

booksmedicos.org

de estos conceptos contribuirá a reducir la incertidumbre del médico acerca del diagnóstico. En otras situaciones, puede que sólo aumente la comprensión del grado de incertidumbre. En ocasiones, puede convencer al médico de que su nivel de incertidumbre es incluso mayor.

En la figura 8-1 se muestra una forma sencilla de contemplar las relaciones entre los resultados de una prueba y el auténtico diagnóstico. La prueba se considera positiva (anormal) o negativa (normal), y la enfermedad está presente o ausente. Por lo tanto, la prueba puede proporcionar cuatro tipos de resultados posibles, dos de los cuales son correctos (verdaderos) y dos, incorrectos (falsos). La prueba ha proporcionado un resultado correcto cuando es positiva en presencia de enfermedad (**verdadero positivo**) o negativa en ausencia de enfermedad (**verdadero negativo**). Por otro lado, la prueba ha proporcionado un resultado erróneo si ha sido positiva cuando la enfermedad está ausente (**falso positivo**) o negativa cuando la enfermedad está presente (**falso negativo**).

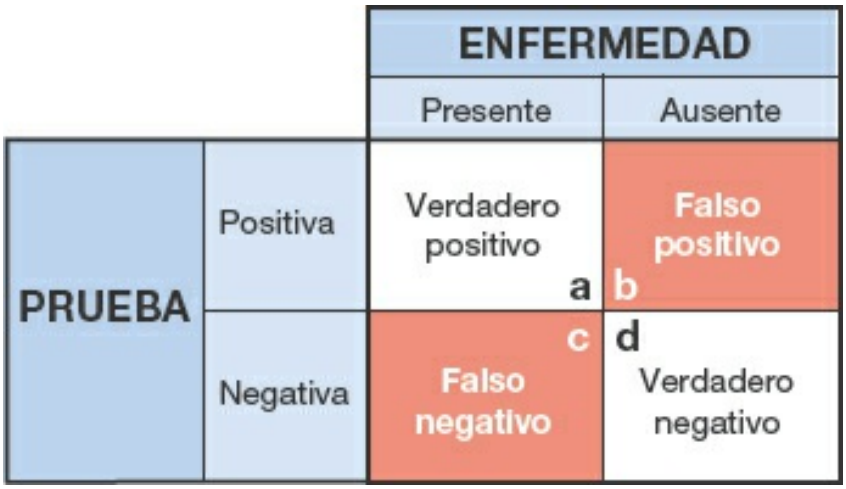


Figura 8-1 Relación entre el resultado de una prueba diagnóstica y la presencia de enfermedad. Existen dos posibilidades de que el resultado de la prueba sea correcto (verdadero positivo y verdadero negativo), y dos posibilidades de que el resultado sea incorrecto (falso positivo y falso negativo).

El estándar de oro

Se considera que la exactitud de una prueba se relaciona con alguna forma de comprobar si la enfermedad está realmente presente o no, es decir, una indicación firme de la verdad, denominada habitualmente **estándar de oro** (también **patrón de referencia** o **patrón de comparación**). En ocasiones, el propio patrón de exactitud es una prueba relativamente sencilla y económica,

booksmedicos.org

como la prueba rápida de antígeno estreptocócico para detectar la presencia de estreptococo del grupo A, y validar así la impresión clínica de amigdalitis estreptocócica, o una prueba de detección de anticuerpos, para confirmar la infección por el virus de la inmunodeficiencia humana. Con mayor frecuencia, para asegurarse de que la enfermedad está presente o ausente, hay que decidir la realización de pruebas relativamente más complejas, caras o con un mayor nivel de riesgo, entre las que se encuentran la biopsia, la exploración quirúrgica, las pruebas de diagnóstico por imagen y, obviamente, la autopsia.

En el caso de las enfermedades que no se resuelven de forma espontánea y que suelen manifestarse después de varios meses o incluso años después de hacer la prueba, los resultados del seguimiento pueden ser válidos como estándar de oro. La detección sistemática de la mayor parte de los cánceres y enfermedades degenerativas crónicas se incluyen dentro de esta categoría. En este caso, puede realizarse la validación incluso si la confirmación del rendimiento de la prueba no resulta factible *in situ*, debido a que el estándar de oro inmediatamente disponible entraña un riesgo demasiado elevado o resulta excesivamente complicado o caro. Si se opta por el seguimiento, la duración del mismo debe ser lo bastante prolongada como para que la enfermedad se manifieste, pero no lo suficiente para que puedan surgir nuevos casos de enfermedad después de las pruebas iniciales (v. cap. 10).

Puesto que emplear formas más exactas de establecer la verdad resulta casi siempre más costoso, más peligroso, o ambas cosas, los médicos y los pacientes prefieren pruebas que sean más sencillas que el estándar de oro estricto, como mínimo, al principio. Para determinar la causa de una neumonía, se utilizan radiografías torácicas y frotis de esputos, en lugar de un estudio de broncoscopia o biopsia pulmonar para examinar el tejido pulmonar enfermo. Para comprobar el diagnóstico de un infarto agudo de miocardio, lo primero que se emplea son electrocardiogramas y análisis de sangre, en lugar de cateterismo o pruebas de diagnóstico por imagen. Las pruebas más sencillas sustituyen a otros métodos más complejos, pero más exactos, para verificar la presencia de enfermedad, y se supone que existe cierto riesgo de clasificación errónea. La seguridad y la comodidad de estas pruebas más sencillas justifican este riesgo, pero las pruebas sencillas sólo resultan útiles cuando se conocen los riesgos de clasificación errónea y se consideran aceptablemente bajos; de lo contrario, es necesaria una firme comparación de su exactitud con un patrón de referencia adecuado.

booksmedicos.org

Falta de información sobre las pruebas negativas

El objetivo de todos los estudios clínicos diseñados para describir el valor de las pruebas diagnósticas debe ser obtener datos para las cuatro casillas que se muestran en la figura 8-1. Sin poseer todos estos datos, no es posible evaluar de forma completa la exactitud de una prueba. La mayor parte de la información sobre el valor de una prueba diagnóstica no se obtiene de los ámbitos de investigación, sino de los ámbitos clínicos, y es en estas circunstancias cuando los médicos utilizan la prueba para la asistencia a los pacientes. Por razones éticas, cuando las pruebas diagnósticas preliminares son negativas, no suele verse justificado proseguir con una evaluación más exhaustiva. Se muestran reacios por naturaleza a iniciar un estudio diagnóstico intensivo, con sus costos y riesgos asociados, salvo que las pruebas preliminares sean positivas. Debido a ello, los datos sobre el número de verdaderos negativos y falsos negativos generados por una prueba (casillas *c* y *d* en la fig. 8-1) tienden a ser mucho menos completos en la bibliografía médica que los datos recogidos sobre las pruebas que han resultado positivos.

Este problema puede surgir en estudios de pruebas de detección sistemática o tamizaje, debido a que las personas con pruebas negativas no suelen someterse a pruebas adicionales, especialmente si estas implican procedimientos cruentos como las biopsias. Un método para superar este problema es utilizar la sangre almacenada o los bancos de tejidos. Una investigación sobre la prueba del antígeno prostático específico (PSA) para la detección del cáncer de próstata examinó este problema examinando sangre almacenada de hombres que posteriormente sufrieron cáncer de próstata y de los que no presentaron dicha enfermedad (2). Los resultados mostraron que, para un nivel de PSA de 4,0 ng/ml, la sensibilidad a lo largo de los 4 años posteriores fue del 73 %, y la especificidad, del 91 %. Los investigadores pudieron rellenar las cuatro casillas sin tener que someter a pruebas adicionales a los pacientes cuyos resultados fueron negativos. (Los conceptos de sensibilidad y especificidad se exponen a continuación.)

Falta de información sobre los resultados de las pruebas en pacientes sanos

Algunos tipos de pruebas son generalmente anómalas en personas sin la enfermedad o sin molestias. Cuando esto es así, el rendimiento de una prueba puede ser completamente erróneo si esta se aplica sólo a los pacientes que

booksmedicos.org

sufren la enfermedad o los síntomas.

Ejemplo

La resonancia magnética (RM) de la columna lumbar se utiliza para la evaluación de pacientes con lumbalgia. En la RM, los discos intervertebrales de muchos pacientes con lumbalgia aparecen herniados y a menudo el dolor se atribuye al hallazgo. Pero, ¿con qué frecuencia se detectan hernias de disco en las personas que no sufren lumbalgia? Varios estudios realizados en sujetos sin antecedente de lumbalgia o ciática han encontrado discos herniados en 22-58 % y discos inflamados en 24-79 % de los sujetos asintomáticos con una edad promedio de los 35 hasta más de los 60 años (3). En otras palabras, las anomalías en los discos vertebrales son habituales y pueden ser un hallazgo incidental en un paciente con lumbalgia.

Falta de patrones objetivos para las enfermedades

Existen algunas enfermedades para las que sencillamente no hay criterios estrictos para su diagnóstico. La angina de pecho es una de ellas. Sus manifestaciones clínicas se describieron hace casi un siglo, y aun así no existe un modo mejor de corroborar su presencia que a través de un minucioso interrogatorio. Es cierto que existe un gran número de fenómenos relacionados con este síndrome clínico que se pueden medir objetivamente, por ejemplo, la presencia de estenosis coronaria en una angiografía, el retraso de la perfusión en una prueba de esfuerzo con talio, y anomalías características en electrocardiogramas realizados tanto en reposo como durante el esfuerzo. Todos ellos se encuentran con mayor frecuencia en pacientes con una presunta angina de pecho, pero ninguno está lo suficientemente vinculado al síndrome clínico como para que pueda utilizarse como referencia a partir del cual se considere la enfermedad como presente o ausente.

Otros ejemplos de afecciones difíciles de diagnosticar por falta de pruebas de referencia sencillas son los sofocos, la enfermedad de Raynaud, el síndrome del colon irritable y el autismo. En un esfuerzo por normalizar la práctica, existen grupos de expertos que a menudo crean listas de síntomas y resultados de otras pruebas que pueden combinarse para diagnosticar la enfermedad clínica. Sin embargo, puesto que no existe un estándar de oro, puede que estas listas no sean del todo correctas. Puede presentarse un razonamiento circular: la validez de una prueba de laboratorio se comprueba

booksmedicos.org

comparando sus resultados con un diagnóstico clínico basado en un minucioso interrogatorio de los síntomas y en una exploración física. Una vez comprobada, la prueba se utiliza para validar el diagnóstico clínico obtenido mediante el interrogatorio y la exploración física.

Consecuencias de los patrones de referencia imperfectos

Debido a estas dificultades, a veces, los médicos en ejercicio no pueden encontrar información sobre el resultado de la comparación entre las pruebas que utilizan y un patrón perfectamente fiable. Como patrón de validez, deben elegir otra prueba que se reconoce como imperfecta, pero que se considera la mejor de las disponibles. Por esta razón, pueden verse forzados a comparar una prueba imperfecta con otra que también lo es, tomando una de ellas como patrón de validez porque se ha utilizado durante más tiempo o porque, por consenso de los expertos, se considera superior. Al hacerlo así puede surgir una paradoja. Si la nueva prueba se compara con una prueba de referencia más antigua (pero imperfecta), la nueva prueba puede parecer peor, aunque en realidad sea mejor. Por ejemplo, si la nueva prueba fuera más sensible que la prueba de referencia, los pacientes adicionales identificados por la nueva prueba se considerarían falsos positivos en relación con la prueba antigua. Del mismo modo, si la prueba resulta negativa con mayor frecuencia en los pacientes que realmente no sufren la enfermedad, los resultados en dichos pacientes se considerarán falsos negativos al compararlos con la prueba antigua. Por lo tanto, el rendimiento de una nueva prueba no puede superar el de una prueba establecida como estándar de oro, ya que la primera parecerá inferior aunque se aproxime más al valor real, salvo que se utilicen estrategias especiales.

Ejemplo

En la detección sistemática del cáncer de colon y los pólipos adenomatosos que pueden ser precursores de cáncer se comparó la colonoscopia con tomografía computarizada (o colonoscopia virtual) con la colonoscopia tradicional (óptica) (4). Ambas pruebas se realizaron de forma que el médico que interpretaba cada una de ellas desconocía los resultados de la otra. La colonoscopia tradicional suele considerarse el estándar de oro para la identificación de pólipos o cáncer de colon en adultos asintomáticos. Sin embargo, la colonoscopia virtual detectó un cáncer de colon y pólipos adenomatosos (especialmente aquellos detrás de los pliegues del colon) que la colonoscopia tradicional había pasado por alto. Para evitar la penalización de la nueva prueba en

booksmedicos.org

comparación con la antigua, los investigadores idearon ingeniosamente un nuevo estándar de oro (la repetición de la colonoscopia óptica una vez revisados los resultados de ambos procedimientos) siempre que hubiera discrepancias entre las pruebas.

SENSIBILIDAD Y ESPECIFICIDAD

La figura 8-2 resume algunas de las relaciones que existen entre una prueba diagnóstica y la presencia real de la enfermedad. Se trata de una ampliación de la figura 8-1, a la que se han añadido algunas definiciones útiles. La mayor parte de lo que queda de capítulo aborda estas relaciones de forma más detallada.

Ejemplo

La figura 8-3 ilustra estas relaciones con un estudio real (5). La trombosis venosa profunda (TVP) de las extremidades inferiores es una enfermedad grave que puede causar una embolia pulmonar; los pacientes con TVP deben tratarse con anticoagulantes. Sin embargo, puesto que la anticoagulación también entraña riesgos, es importante distinguir entre los pacientes con y sin TVP. La ecografía de compresión presenta una gran sensibilidad y especificidad para la trombosis proximal, y se ha utilizado para confirmar o descartar la TVP. La ecografía de compresión resulta cara, y necesita personal muy calificado, por lo que se empezó a buscar una prueba diagnóstica más sencilla. La medición de los marcadores de fibrinólisis endógena, o estudio de dímeros D, se perfeccionó hasta que se pudo realizar a partir de análisis de sangre y se evaluó en el diagnóstico de la TVP. La figura 8-3 muestra el rendimiento de un estudio de dímeros D en el diagnóstico de la TVP. El estándar de oro del estudio fue el resultado de la ecografía de compresión, un seguimiento de tres meses o ambas cosas.

booksmedicos.org

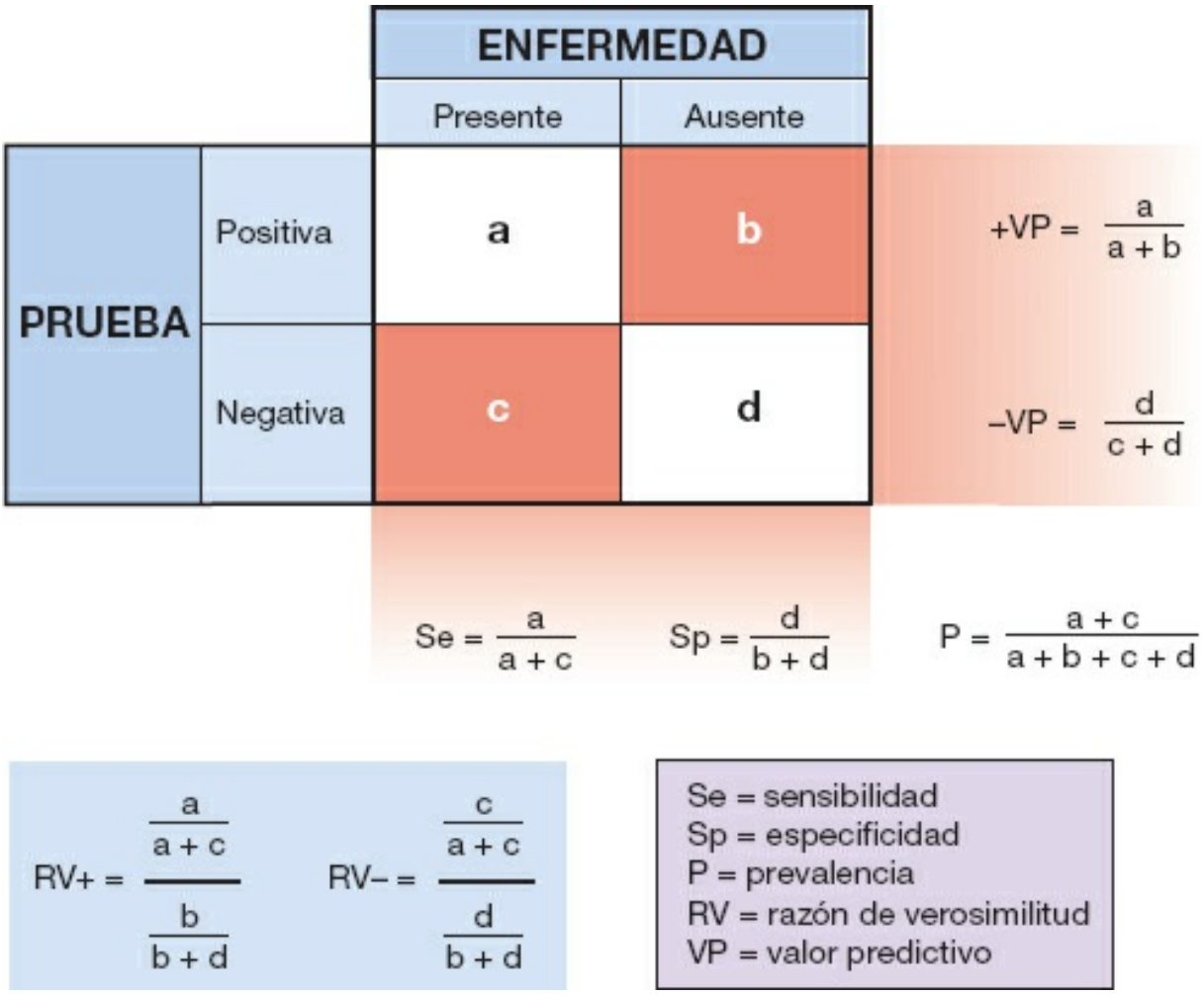


Figura 8-2 ■ Definiciones y características de prueba diagnóstica. Obsérvese que los cálculos para RV+ corresponden a los mismos que para Se/(1 – Sp), y que los cálculos para RV– están dados por los mismos que para (1 – Se)/Sp.

booksmedicos.org

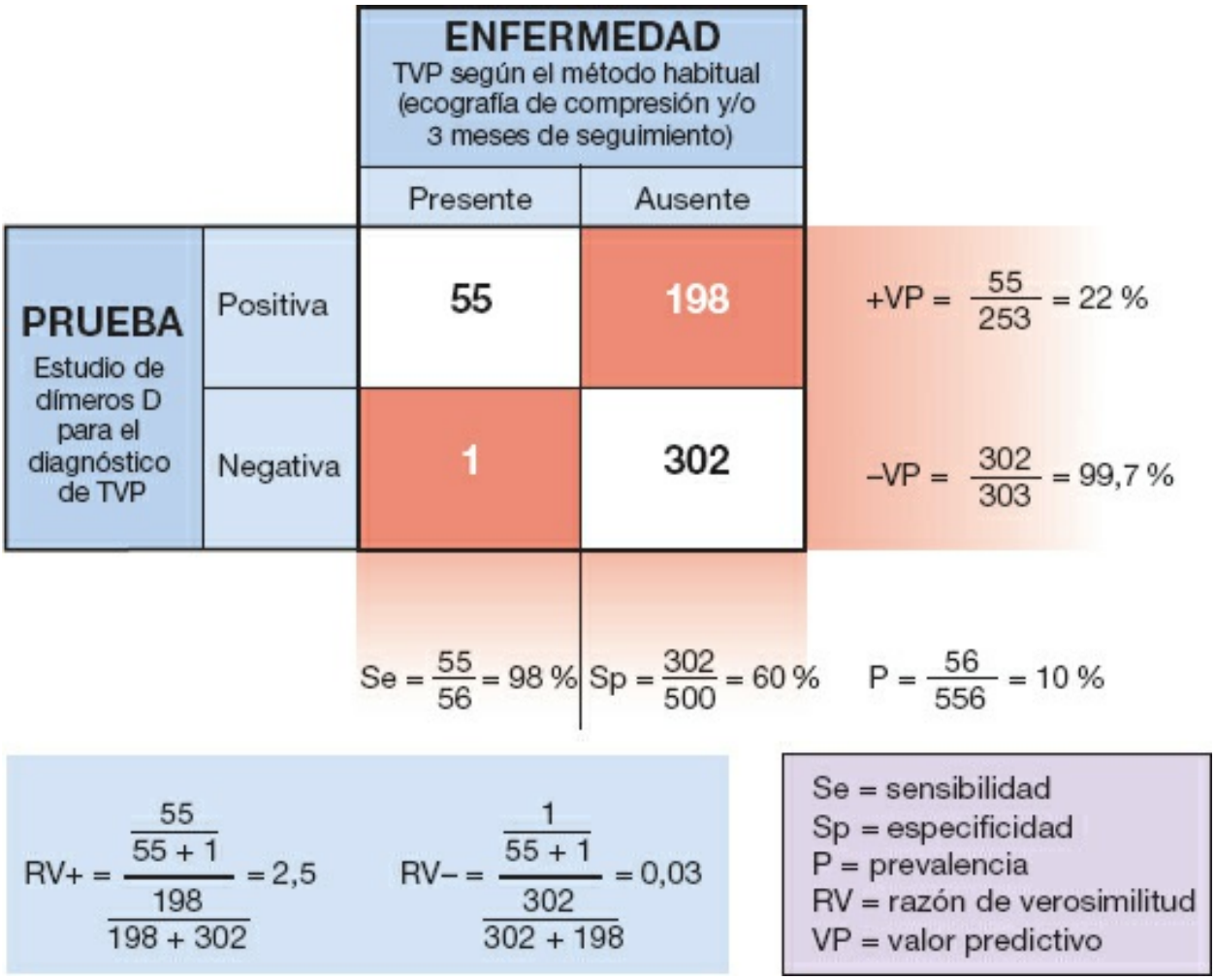


Figura 8-3 ■ Características diagnósticas de un estudio de dímeros D en el diagnóstico de la trombosis venosa profunda (TVP). (Datos de Bates SM, Kearon C, Crowther M, et al. A diagnostic strategy involving a quantitative latex D-dimer assay reliably excludes deep venous thrombosis. Ann Intern Med 2003;138:787-794.)

Definiciones

Tal como se observa en la figura 8-2, la **sensibilidad** se define como la proporción de personas con una enfermedad y con un resultado positivo en una prueba para detectar dicha enfermedad. Una prueba sensible casi nunca pasará por alto a pacientes que sufren la enfermedad. La **especificidad** es la proporción de personas sin la enfermedad y con un resultado negativo en la prueba realizada para detectarla. Una prueba específica rara vez clasificará como enfermas a personas que realmente no sufren la enfermedad.

Si aplicamos estas definiciones al ejemplo de la TVP (fig. 8-3), observamos que 55 de 56 pacientes con TVP obtuvieron resultados positivos en la prueba de dímeros D (una sensibilidad del 98 %). Sin embargo, de los

booksmedicos.org

500 pacientes que no sufrían TVP, sólo 302 obtuvieron resultados negativos correctos en la prueba de dímeros D (una especificidad del 60 %).

Utilización de pruebas sensibles

Los médicos deben tener en cuenta la sensibilidad y la especificidad de una prueba diagnóstica cuando vayan a optar por ella. Debe elegirse una prueba sensible (es decir, una prueba que sea habitualmente positiva en presencia de la enfermedad) cuando un resultado erróneo conlleve un daño grave. Esto prevalece, por ejemplo, cuando existe alguna razón para sospechar una enfermedad peligrosa pero tratable, como la tuberculosis, la sífilis o la enfermedad de Hodgkin, linfomas o en un paciente en el que se sospecha una TVP. Las pruebas sensibles también son útiles en las primeras etapas de un estudio diagnóstico bien elaborado, cuando existe un buen número de posibilidades que considerar, para reducir el número de dichas posibilidades. En estas situaciones, las pruebas diagnósticas se utilizan para descartar enfermedades con un resultado negativo en una prueba de gran sensibilidad (como en el ejemplo de la TVP). Por poner otro ejemplo, en determinados pacientes, en las primeras etapas de evaluación de unos infiltrados pulmonares y pérdida de peso, podría optarse por la prueba de anticuerpos del VIH, de gran sensibilidad, para descartar una infección relacionada con el SIDA. En resumen, una prueba con una gran sensibilidad es más útil para el médico cuando su resultado es negativo.

Utilización de pruebas específicas

Las pruebas específicas son útiles para confirmar (o «aceptar») un diagnóstico sugerido por otros datos. Esto se debe a que una prueba con una gran especificidad rara vez será positiva si no existe la enfermedad, es decir, proporcionará pocos resultados falsos positivos. (Obsérvese que, en el ejemplo de la TVP, la prueba de dímeros D no fue lo bastante específica [60 %] para iniciar un tratamiento una vez obtenido el resultado positivo. Todos los pacientes con resultados positivos fueron sometidos a una ecografía de compresión, una prueba mucho más específica.) Las pruebas muy específicas son especialmente necesarias cuando los resultados falsos positivos pueden perjudicar al paciente de un modo físico, emocional o económico. Por este motivo, antes de someter a los pacientes a quimioterapia antineoplásica (con todos sus riesgos asociados, traumas emocionales y costos económicos),

booksmedicos.org

suele precisarse un diagnóstico histológico (una prueba de gran especificidad). En resumen, una prueba muy específica es más útil cuando su resultado es positivo.

Equilibrio entre sensibilidad y especificidad

Obviamente, es deseable contar con una prueba dotada de gran sensibilidad y especificidad. Por desgracia, no suele ser posible que concurran ambas condiciones y, en lugar de ello, siempre que los datos clínicos adquieran un intervalo de valores, existe un equilibrio entre la sensibilidad y la especificidad de una prueba diagnóstica dada. Esto es así siempre que los datos clínicos adopten todo un intervalo de valores. En estas situaciones, la localización de un **punto de corte**, un punto dentro del intervalo continuo que separe lo normal de lo anormal, es fruto de una decisión arbitraria. Como consecuencia, para cualquier resultado expresado en una escala continua, una característica, como la sensibilidad, sólo puede incrementarse en detrimento de otra característica (p. ej., la especificidad). La tabla 8-1 demuestra esta interrelación con respecto al uso de los niveles de péptido natriurético tipo B (PNB) en el diagnóstico de insuficiencia cardíaca congestiva (ICC) entre los pacientes con disnea aguda que acuden a urgencias (6). Si el punto de corte de la prueba se estableciera en una cifra demasiado baja (≥ 50 pg/ml), la sensibilidad sería elevada (97 %), pero como contrapartida la prueba ofrecería una baja especificidad (62 %), lo que requeriría que muchos pacientes con ICC se sometieran a pruebas adicionales. Por otro lado, si el punto de corte se estableciera en una cifra demasiado alta (≥ 150 pg/ml), se pasaría por alto a muchos pacientes con ICC. Los autores sugirieron que una solución aceptable sería un punto de corte de 100 pg/ml, lo que proporcionaría una sensibilidad del 90 % y una especificidad del 76 %. No hay modo, utilizando únicamente una prueba de PNB, de aumentar la sensibilidad y la especificidad de la prueba al mismo tiempo.

Tabla 8-1

Equilibrio entre sensibilidad y especificidad al utilizar los niveles de péptido natriurético tipo B para detectar insuficiencia cardíaca congestiva

Nivel de PNB (pg/ml)	Sensibilidad (%)	Especificidad (%)
50	97	62

booksmedicos.org

80	93	74
100	90	76
125	87	79
150	85	83

PNB, péptido natriurético tipo B.
Adaptado con autorización de Maisel AS, Krishnaswamy P, Nowak RM, et al. Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. N Engl J Med 2002;347: 161-167.

booksmedicos.org

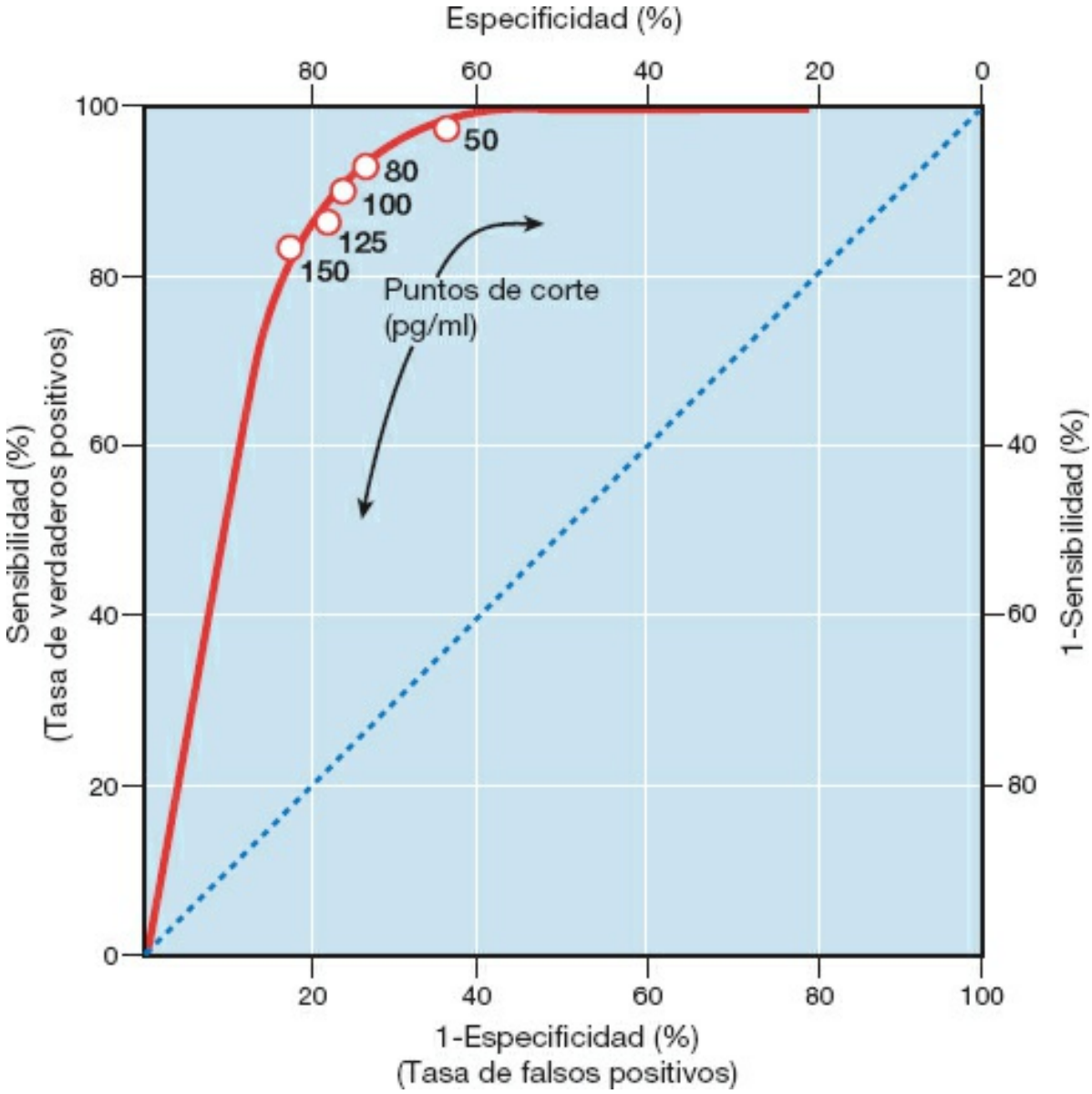


Figura 8-4 ■ Curva ROC (curva de eficacia diagnóstica). La exactitud del péptido natriurético tipo B (PNB) en el diagnóstico urgente de insuficiencia cardíaca con diversos puntos de corte del PNB entre la disnea por insuficiencia cardíaca congestiva y otras causas. (Adaptado con autorización de Maisel AS, Krishnaswamy P, Nowak RM, et al. Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. N Engl J Med 2002;347:161-167.)

La curva ROC

Otra forma de expresar la relación entre la sensibilidad y la especificidad de una prueba concreta consiste en elaborar una **curva de eficacia diagnóstica,** o **curva ROC** *(receiver operator characteristic)*. En la figura 8-4 se muestra una curva ROC para los niveles de PNB de la tabla 8-1. Se ha elaborado representando la tasa de verdaderos positivos (sensibilidad) frente a la tasa de

booksmedicos.org

falsos positivos (1– especificidad) a lo largo de un intervalo de valores de corte. Los valores en los ejes discurren entre una probabilidad de 0 y 1 (o, de forma alternativa, del 0-100 %). La figura 8-4 ilustra el equilibrio entre sensibilidad y especificidad.

Las pruebas que consiguen una buena distinción se agrupan en el extremo superior izquierdo de la curva ROC; en ellas, conforme aumenta la sensibilidad (a medida que el punto de corte se reduce), no se pierde (o se pierde muy poca) especificidad, hasta que se alcanzan valores de sensibilidad muy elevados. Las pruebas con peor rendimiento presentan curvas más cercanas a la diagonal que discurre entre el extremo inferior izquierdo y el extremo superior derecho. La diagonal muestra la relación entre las tasas de verdaderos positivos y falsos positivos de una prueba sin utilidad, una que no aportará ninguna información adicional a lo que ya se conocía antes de realizar la prueba, lo que equivale a hacer un diagnóstico lanzando una moneda al aire.

booksmedicos.org

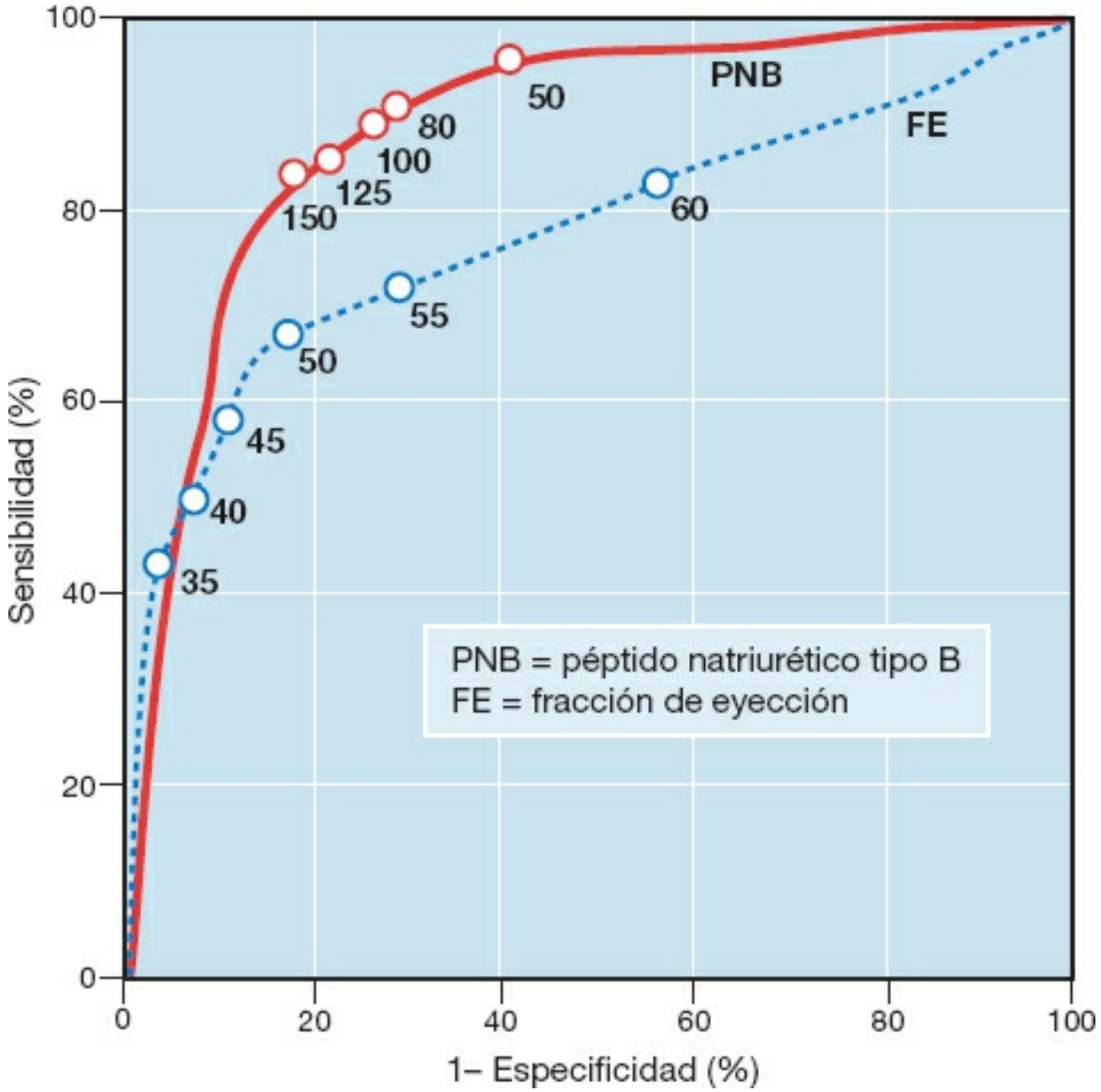


Figura 8-5 ■ Curvas ROC para el PNB y las fracciones de eyección del ventrículo izquierdo por ecocardiografía en el diagnóstico de urgencia de insuficiencia cardíaca congestiva en los pacientes que tienen disnea aguda. En general, el PNB es más sensible y específico que las fracciones de eyección, lo que origina una mayor área bajo la curva. (Redibujado con autorización de Steg PG, Joubin L, McCord J, et al. B-type natriuretic peptide and echocardiographic determination of ejection fraction in the diagnosis of congestive heart failure in patients with acute dyspnea. Chest 2005;128:21-29.)

La curva ROC muestra la importancia del equilibrio entre sensibilidad y especificidad para una prueba concreta, lo que se puede utilizar para decidir dónde debe colocarse el punto de corte óptimo. En general, dicho punto se sitúa en la parte curva (o cerca de ella) de la curva ROC, salvo que existan razones clínicas para reducir al mínimo los falsos negativos o los falsos positivos.

booksmedicos.org

Las curvas ROC representan una forma especialmente valiosa de comparar pruebas alternativas para un mismo diagnóstico. La exactitud global de una prueba puede describirse como el área bajo la curva ROC; cuanto mayor sea el área, mejor será la prueba. La figura 8-5 compara la curva ROC del PNB con una prueba más antigua, la fracción de eyección ventricular determinada por electrocardiografía (7). El PNB es más sensible y específico, con un gran área bajo la curva (0,89) en comparación con la fracción de eyección (0,78). También es más fácil y rápido obtenerlo durante una urgencia, las características de la prueba son importantes en situaciones clínicas cuando se requieren resultados rápidos.

Obviamente, las pruebas más solicitadas son las que ofrecen sensibilidad y especificidad a la vez, además de que pueden ser de gran valor. Sin embargo, los médicos deben trabajar con frecuencia con pruebas cuya sensibilidad y especificidad no son demasiado elevadas. En estos casos, deben utilizar otros medios para sortear el equilibrio entre sensibilidad y especificidad. La forma más habitual de hacerlo es utilizar los resultados de varias pruebas, tal como se explica en este capítulo.

ESTABLECER LA SENSIBILIDAD Y LA ESPECIFICIDAD

A menudo, cuando se presenta una nueva prueba diagnóstica, esta se describe con términos entusiastas, si bien, una vez se ha adquirido la suficiente experiencia con ella, acaba por considerarse deficiente. El entusiasmo inicial, seguido por la decepción, no surge de la deshonestidad de los investigadores, al principio, ni de un escepticismo injustificado por parte de la comunidad médica, al final. Al contrario, este hecho se debe, en primer lugar, a las limitaciones de los métodos utilizados para comprobar las propiedades de la prueba. Es posible que la sensibilidad y la especificidad se describan de forma inexacta porque se ha elegido un estándar de oro inadecuado, como ya se ha comentado anteriormente en este capítulo. Además, existen otros dos aspectos relacionados con la selección de los pacientes con y sin la enfermedad que pueden afectar profundamente, y por igual, a la determinación de la sensibilidad y la especificidad: el **espectro** de pacientes a los que se aplica la prueba y el **sesgo** a la hora de juzgar el rendimiento de la misma. La incertidumbre estadística, relacionada con el hecho de estudiar un número relativamente pequeño de pacientes, también puede dar lugar a estimaciones de sensibilidad y especificidad inexactas.

booksmedicos.org

Espectro de pacientes

Cuando los pacientes utilizados para describir las propiedades de la prueba son diferentes de aquellos a los que se aplicará la prueba en la práctica clínica, pueden surgir dificultades. Con frecuencia, los informes iniciales determinan el valor de la prueba tras comparar a personas claramente enfermas con personas que claramente no lo están, como estudiantes de medicina voluntarios. Puede que la prueba sea capaz de distinguir perfectamente entre estos dos extremos, pero que su rendimiento sea deficiente cuando las diferencias son sutiles. Asimismo, los pacientes con la enfermedad a menudo difieren en cuanto a la duración, el estadio o la gravedad de la misma, y la sensibilidad de la prueba tenderá a ser mayor en los pacientes que revisten mayor gravedad.

Ejemplo

El cáncer de ovario, el cuarto cáncer no cutáneo más habitual en las mujeres, se ha extendido más allá del ovario en el momento de su detección clínica en la mayoría de las pacientes, y las tasas de supervivencia a 5 años para estas pacientes son de un 30 %. Debido a que no existe una prueba efectiva de tamizaje para el cáncer de ovario, se desarrollaron guías clínicas para la referencia temprana de las pacientes con masas pélvicas a los especialistas en ginecología y obstetricia, con la esperanza de detectar esta enfermedad en una etapa susceptible de ser curada. Se realizó un estudio sobre la precisión de estas directrices en pacientes evaluadas por una masa pélvica (8). En las postmenopáusicas, la sensibilidad global de las directrices fue del 93 % y la especificidad del 60 %. Sin embargo, cuando se examinó de acuerdo con la etapa del cáncer, la sensibilidad para la etapa temprana fue menor que para la etapa tardía, 80 % frente al 98 %. Diez de los 14 casos de cáncer de ovario primario que se pasaron por alto al usar las directrices correspondían a la etapa I, la más temprana. Por lo tanto, la sensibilidad de estas guías para detectar este cáncer dependía de la mezcla particular de etapas en los pacientes enfermos utilizados para describir las características de la prueba. Irónicamente, la sensibilidad de las guías fue más baja para la etapa que tendría mayor utilidad clínica.

Algunas personas en las que se sospecha la enfermedad pueden tener otras afecciones que produzcan un resultado positivo, lo que provoca una tasa de falsos positivos más elevada y una especificidad menor. En el ejemplo de las guías clínicas para la evaluación del cáncer de ovario, la especificidad fue baja para todas las etapas de cáncer (60 %). Una razón para ello es que los niveles del marcador tumoral CA-125, recomendado por las directrices, están elevados en muchas otras enfermedades y trastornos además del cáncer de

booksmedicos.org

ovario. Estas condiciones extrañas disminuyeron la especificidad y aumentaron la tasa de falsos positivos de las directrices. La baja especificidad de las pruebas diagnósticas y de tamizaje es un problema importante en el cáncer de ovario y puede hacer que muchas mujeres sin cáncer sean sometidas a cirugía.

El espectro y la prevalencia de la enfermedad son especialmente importantes cuando una prueba se utiliza para el tamizaje o detección temprana en lugar de para el diagnóstico (v. cap. 10 para una discusión más detallada sobre el tamizaje). En teoría, la sensibilidad y la especificidad de una prueba no dependen de la prevalencia de pacientes enfermos incluidos en la muestra en la que se evalúa dicha prueba. (Inténtelo usted mismo con la fig. 8-2 para confirmarlo.) Sin embargo, en la práctica, algunas características de los pacientes, como el estadio y la gravedad de la enfermedad, pueden estar relacionadas tanto con la sensibilidad y la especificidad de la prueba, como con la prevalencia, debido a que diferentes tipos de pacientes se encuentran en situaciones de prevalencia elevada y baja. El tamizaje de una prueba ilustra este punto. Esta detección o tamizaje implica la utilización de una prueba en una población asintomática en la que la prevalencia de la enfermedad es generalmente baja y el espectro de la enfermedad favorece el reconocimiento de casos precoces y menos graves. En estas situaciones, la sensibilidad tiende a ser inferior y la especificidad mayor que cuando se aplica la misma prueba a pacientes en los que se sospecha la enfermedad, ya que un mayor número de ellos la sufre en un estadio más avanzado.

Ejemplo

Se realizó un estudio sobre la sensibilidad y la especificidad de la exploración clínica de la mama en cerca de 750 000 mujeres (9). Cuando esta exploración clínica se utilizó como prueba *diagnóstica* en mujeres con síntomas mamarios, la sensibilidad y la especificidad del cáncer de mama fue del 85 % y el 73 %, respectivamente. Sin embargo, cuando la exploración se utilizó como prueba de *tamizaje* en mujeres que estaban asintomáticas, la sensibilidad descendió hasta el 36 % y la especificidad aumentó hasta el 96 %.

Sesgos

La sensibilidad y la especificidad de una prueba deben establecerse de forma independiente de los medios utilizados para llegar al diagnóstico real. De lo

booksmedicos.org

contrario, podría incurrirse en una evaluación sesgada de las propiedades de la prueba. Como ya se ha señalado, si la prueba se evalúa utilizando datos obtenidos durante el curso de la evaluación clínica de los pacientes en los que se sospecha la enfermedad, un resultado positivo en la prueba puede impulsar al médico a seguir buscando un diagnóstico, lo que haría aumentar las probabilidades de descubrir la enfermedad. Por otro lado, un resultado negativo puede hacer que el médico desestime la necesidad de realizar pruebas adicionales, lo que aumentaría las posibilidades de que la enfermedad, en caso de existir, se pasara por alto.

Por lo tanto, cuando se evalúan la sensibilidad y la especificidad de una prueba, el resultado de esta no debe formar parte de la información utilizada para establecer el diagnóstico. Al estudiar el diagnóstico de TVP con el análisis de dímeros D (comentado previamente), los investigadores se aseguraron que los médicos que llevaban a cabo las pruebas de referencia (ecografía y evaluación de seguimiento) no conocieran los resultados de las mediciones de dímeros D, de forma que dichos resultados no pudieran influir (sesgos) en sus interpretaciones de la ecografía (10).

En la asistencia clínica sistemática, este tipo de sesgo puede utilizarse en beneficio propio, especialmente si el resultado de la prueba tiene una interpretación subjetiva. Muchas interpretaciones de radiografías son subjetivas, y es fácil que se vean influidas por la información clínica suministrada. Todos los médicos han tenido que interpretar de nuevo las radiografías debido a una impresión clínica o, a la inversa, han vuelto a examinar radiografías antiguas en las que se pasó por alto un hallazgo debido a que en su momento no se comunicó un hecho clínico determinado, y por ello no se prestó atención a una región concreta de la radiografía. Tanto para reducir al mínimo estos sesgos como para utilizarlos en beneficio propio, algunos radiólogos prefieren examinar las radiografías dos veces, primero sin disponer de la información clínica y, posteriormente, en función de esta.

booksmedicos.org

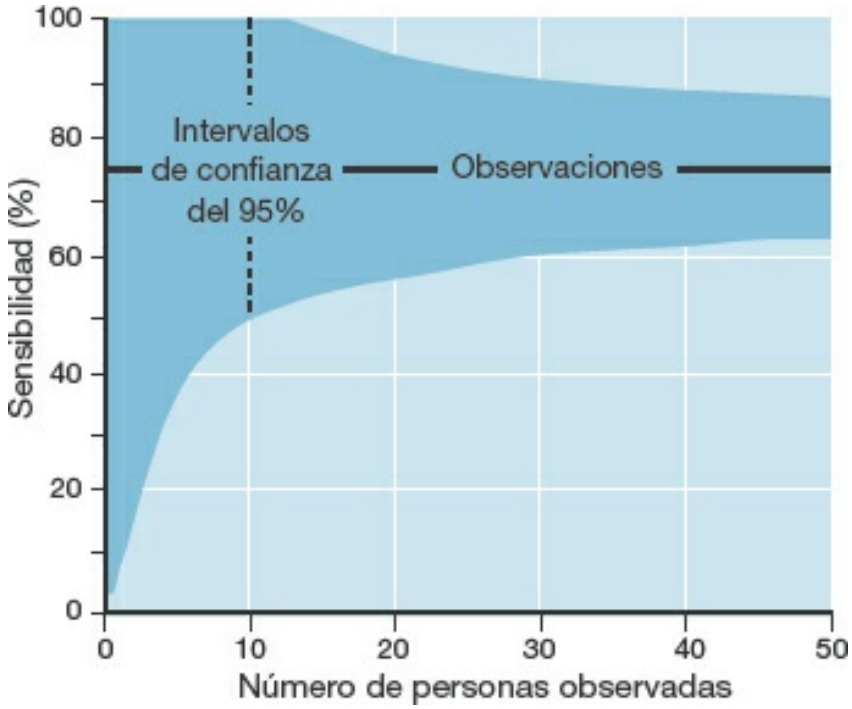


Figura 8-6 ■ Precisión de una estimación de la sensibilidad. Intervalo de confianza del 95 % para una sensibilidad observada del 75 % en función del número de personas observadas.

Todos los sesgos comentados tienden a incrementar el grado de acuerdo entre la prueba y el estándar de oro. Es decir, tienden a hacer que la prueba parezca más exacta de lo que realmente es.

Azar

Habitualmente, los valores para la sensibilidad y la especificidad se calculan a partir de observaciones realizadas en muestras relativamente pequeñas de personas con y sin la enfermedad objeto de interés. Debido al azar (variación aleatoria) que afecta a cualquier muestra, especialmente si esta es de pequeño tamaño, la sensibilidad y la especificidad reales de la prueba pueden estar representadas de forma errónea incluso si en el estudio se ha evitado el sesgo. Los valores concretos observados son compatibles con un intervalo de valores reales, habitualmente caracterizados por los «intervalos de confianza del 95 %» (v. cap. 12)†. La amplitud de este intervalo de valores define la precisión de las estimaciones de sensibilidad y especificidad. Por consiguiente, los valores de sensibilidad y especificidad notificados no deben tomarse de forma demasiado literal si sólo se ha estudiado a un número reducido de pacientes.

La figura 8-6 muestra cómo aumenta la precisión de las estimaciones de

booksmedicos.org

la sensibilidad, a medida que lo hace el número de personas sobre el que se basan dichas estimaciones. En este ejemplo concreto, la sensibilidad observada de la prueba diagnóstica es del 75 %. También muestra que si esta estimación se basa únicamente en 10 pacientes, sólo por efecto del azar la sensibilidad real podría descender hasta el 45 % o aumentar hasta el 100 %. Cuando se estudia un número mayor de pacientes, el intervalo de confianza del 95 % se estrecha y aumenta la precisión de las estimaciones.

[†]El intervalo de confianza del 95 % de una proporción puede calcularse fácilmente mediante la siguiente fórmula, basada en el teorema binomial:

$$p \pm 2\sqrt{\frac{p(1-p)}{N}}$$

donde *p* representa la proporción observada, y *N*, el número de personas observadas. Para ser más exactos, multiplicar por 1,96 en lugar de 2.

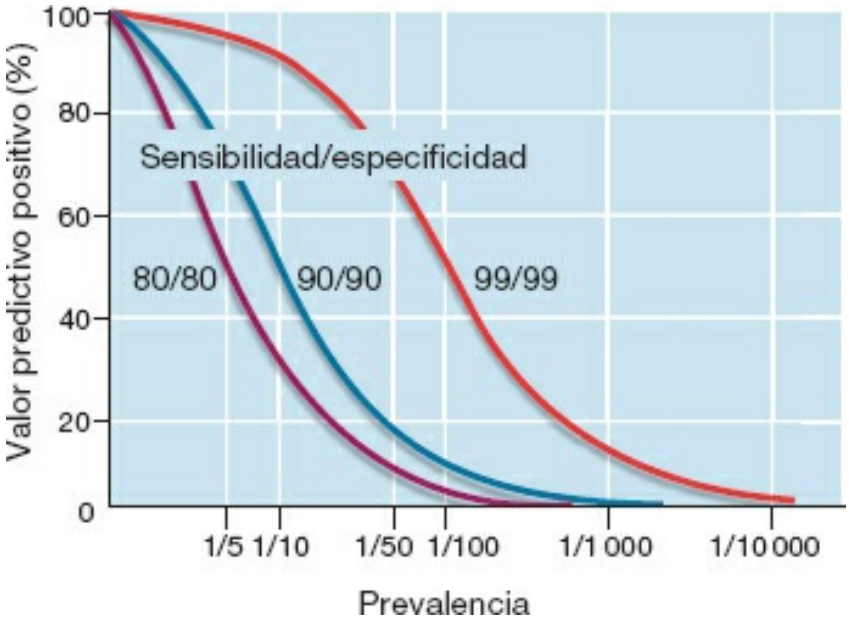


Figura 8-7 ■ Valor predictivo positivo en función de la sensibilidad, especificidad y prevalencia de enfermedad.

VALOR PREDICTIVO

La sensibilidad y la especificidad son propiedades de una prueba que deben tenerse en cuenta a la hora de decidir si se utiliza dicha prueba. Sin embargo, una vez que están disponibles los resultados de una prueba diagnóstica, ya

booksmedicos.org

sean estos positivos o negativos, la sensibilidad y la especificidad de la prueba dejan de ser relevantes, ya que estos valores han sido obtenidos por personas en las que la presencia o ausencia de la enfermedad ya se conoce. Si se conociera el estado de enfermedad del paciente, no sería necesario solicitar la prueba. Para el médico, el dilema consiste en determinar si el paciente sufre o no la enfermedad. (De hecho, los médicos suelen estar más preocupados por esta cuestión que por la sensibilidad y la especificidad de la prueba.)

Definiciones

Dados los resultados de una prueba, la probabilidad de enfermedad se denomina **valor predictivo** de la prueba (v. fig. 8-2). El **valor predictivo positivo** es la probabilidad de una enfermedad en un paciente con un resultado positivo (anormal) en una prueba. El **valor predictivo negativo** es la probabilidad de *no* sufrir la enfermedad cuando el resultado de la prueba ha sido negativo (normal) El valor predictivo responde la pregunta: «Si mi paciente ha tenido un resultado positivo/negativo en la prueba, ¿qué probabilidad hay de que sufra/no sufra la enfermedad?» En ocasiones al valor predictivo también se le llama **probabilidad posterior** (o **posprueba)**, es decir, la probabilidad de enfermedad *tras* conocerse el resultado. La figura 8-3 ilustra estos conceptos. De los 253 pacientes con un resultado positivo en los estudios de dímeros D, sólo el 55 (22 %) sufría TVP (valor predictivo positivo). El valor predictivo negativo de la prueba fue mucho más elevado, casi del 100 %.

A veces, se emplea el término **exactitud** para resumir el valor global de una prueba. La exactitud es la proporción de todos los resultados de la prueba, tanto positivos como negativos, que son correctos. En el ejemplo de la TVP de la figura 8-3, la exactitud de los estudios de dímeros D fue del 64 %. (Calcúlelo usted mismo.) El área bajo la curva ROC es otra medida de resumen útil de la información proporcionada por el resultado de una prueba. Sin embargo, estas medidas de resumen suelen ser demasiado elementales para que resulten útiles en la práctica clínica, debido a que la información específica de sus partes integrantes (sensibilidad, especificidad y valor predictivo en puntos de corte específicos) se pierde cuando se expresan en forma de un solo número.

Determinantes del valor predictivo

booksmedicos.org

El valor predictivo de una prueba no es una propiedad únicamente de la prueba, pues está determinado por la sensibilidad y la especificidad de la misma, así como por la prevalencia de la enfermedad en la población objeto de estudio, cuando el término **prevalencia** tiene su significado habitual: la proporción de personas en una población definida en un punto concreto del tiempo en que sufren la enfermedad. Otro significado de prevalencia es el de **probabilidad previa** (o **preprueba**), es decir, la probabilidad de una enfermedad antes de que se conozca el resultado de la prueba. (En el cap. 2 se aborda la prevalencia de forma exhaustiva.)

Cuanto más sensible sea una prueba, mayor será su valor predictivo negativo (mayor seguridad tendrá el médico de que un resultado negativo descarta la enfermedad que se está buscando). A la inversa, cuanto más específica sea una prueba, mejor será su valor predictivo positivo (mayor seguridad tendrá el médico de que un resultado positivo confirma el diagnóstico que se está buscando). Debido a que el valor predictivo también está influido por la prevalencia, no es independiente del entorno en el que se utiliza la prueba. Los resultados positivos, incluso de una prueba muy específica, si proceden de pacientes con una escasa probabilidad de sufrir la enfermedad, serán en gran medida falsos positivos. De forma similar, los resultados negativos, incluso de una prueba muy sensible, si proceden de pacientes con una elevada probabilidad de sufrir la enfermedad, serán posiblemente falsos negativos. En resumen, la interpretación del resultado de una prueba diagnóstica, ya sea positivo o negativo, varía de un entorno a otro, en función de la prevalencia de la enfermedad en el entorno concreto.

El motivo de por qué la prevalencia debe afectar a la interpretación de un resultado no resulta intuitivamente obvio. Para los escépticos, puede ser útil considerar el rendimiento de una prueba en cada extremo de la prevalencia. Hay que recordar que no importa el grado de sensibilidad o especificidad de una prueba (aunque sea casi perfecta); siempre habrá una pequeña proporción de pacientes a los que clasificará de forma errónea. Imagínese una población en la que nadie sufre la enfermedad. En un grupo como este, todos los resultados positivos, aunque sean de una prueba muy específica, serán falsos positivos. Por lo tanto, conforme la prevalencia de una enfermedad en una población se aproxima a cero, el valor predictivo positivo de la prueba también se aproximará a cero. A la inversa, si todos los pacientes de una población sometida a la prueba padecen la enfermedad, todos los resultados negativos serán falsos negativos, aunque procedan de una prueba muy

booksmedicos.org

sensible. Conforme la prevalencia se aproxima al 100 %, el valor predictivo negativo se aproxima a cero. Otra forma de que los escépticos se convenzan de la existencia de estas relaciones es que prueben con la figura 8-3, que mantengan la sensibilidad y la especificidad constantes, cambien la prevalencia y calculen los valores predictivos resultantes.

En la figura 8-7 se ilustra el efecto de la prevalencia sobre el valor predictivo positivo, para una prueba con niveles de sensibilidad y especificidad diferentes (aunque, por lo general, elevados). Cuando la prevalencia de la enfermedad en la población sometida a la prueba es relativamente alta (por encima de determinados porcentajes), la prueba funciona bien. Pero con prevalencias inferiores, el valor predictivo positivo puede descender casi hasta cero (el valor predictivo negativo mejora), con lo que la prueba pierde prácticamente su utilidad. (La figura ilustra por qué el valor predictivo positivo de una prueba es mucho mejor para los estudios diagnósticos, que a menudo evalúan la prueba con unos pocos cientos de pacientes, en comparación con los estudios de tamizaje, que por lo general deben usar decenas de miles de pacientes, debido a la diferencia subyacente en la prevalencia de la enfermedad en las dos situaciones). A veces es posible manipular la prevalencia de la enfermedad en las situaciones clínicas, a fin de que una prueba diagnóstica se vuelva más útil.

Conforme se reducen la sensibilidad y la especificidad, la influencia de la prevalencia sobre el valor predictivo es cada vez más pronunciada.

Ejemplo

Como se indicó previamente en este capítulo, la TVP, una causa potencialmente peligrosa de dolor en la pierna, es difícil de diagnosticar sin pruebas especializadas. Muchos pacientes con dolor en la pierna no tienen TVP y es importante diferenciar a los pacientes que tienen TVP de los que no, pues el tratamiento para este trastorno (anticoagulación) es arriesgado. Podría ser útil contar con una prueba diagnóstica rápida y sencilla para descartar TVP, pero, como se deduce de la figura 8-3, las pruebas de dímeros D son sensibles pero no específicas. Sin embargo, incluso una prueba con relativamente poca especificidad puede funcionar relativamente bien en un grupo de pacientes con baja prevalencia de la enfermedad. Se realizó un esfuerzo para descartar la TVP y mejorar el valor predictivo negativo al clasificar a los pacientes en grupos con diferentes probabilidades (prevalencias) de TVP antes de realizar la prueba diagnóstica. En una síntesis de 14 estudios que evaluaron el uso de pruebas de dímeros D en más de 8 000 pacientes, los investigadores encontraron que la prevalencia global de TVP fue del 19 % (11). Utilizando una simple regla clínica validada que comprende varios hallazgos clínicos y elementos de la historia, se clasificó a los pacientes en grupos de baja (5 %), intermedia (17 %) y alta (53 %) probabilidad (prevalencia)

booksmedicos.org

de TVP, a quienes luego se les realizó una prueba de dímeros D de alta sensibilidad. La sensibilidad fue alta (95-98 %) en todos los grupos, pero la especificidad varió de acuerdo con la prevalencia de la TVP: 58 % entre los pacientes con probabilidad baja, 41 % entre los pacientes con probabilidad intermedia y 36 % entre los pacientes con probabilidad alta. Aunque una especificidad del 58 % no es muy alta, cuando los resultados de la prueba de dímero D fueron negativos entre los pacientes con una probabilidad clínica baja de TVP, el valor predictivo negativo fue > 99 %, menos del 1 % de los pacientes en este grupo se diagnosticaron con TVP al menos 3 meses después del seguimiento. Por lo tanto, la prueba fue especialmente útil para excluir la TVP en los pacientes con una probabilidad baja de padecerla (un 40 % de todos los pacientes) y no hubo necesidad de realizar otras pruebas. Este es un ejemplo de cómo el resultado negativo de una prueba no específica se vuelve clínicamente útil cuando se utiliza en un grupo de pacientes con una prevalencia muy baja de la enfermedad.

Como puede deducirse del ejemplo y la figura 8-6, la prevalencia suele ser más importante que la sensibilidad y la especificidad para determinar el valor predictivo. Una razón para ello es que la prevalencia puede variar a lo largo de un intervalo mucho más amplio que la sensibilidad y la especificidad. La prevalencia de una enfermedad puede variar de 1 en 1 millón a < 1 en 10, en función de la edad, el sexo, los factores de riesgo y los hallazgos clínicos del paciente. Considérese la diferencia entre la prevalencia de hepatopatía en adultos jóvenes y sanos que no consumen drogas, no son promiscuos y consumen alcohol sólo de forma ocasional, y la prevalencia de la misma afección en usuarios de drogas intravenosas con ictericia y numerosas parejas sexuales. Por el contrario, la sensibilidad y la especificidad de las pruebas diagnósticas suelen variar a lo largo de un intervalo mucho más reducido, entre el 50 % y el 99 %.

Estimación de la prevalencia (probabilidad preprueba)

Debido a que la prevalencia de la enfermedad es un determinante tan importante del grado de utilidad de una prueba diagnóstica, los médicos deberían considerar la probabilidad de que exista una enfermedad antes de solicitar una prueba. ¿Cómo puede el médico calcular la prevalencia o probabilidad de una enfermedad concreta en sus pacientes? Hasta hace muy poco, se basaban en depender de las observaciones clínicas y en su experiencia para estimar (en general de forma implícita) la probabilidad preprueba de la mayoría de las enfermedades. Los estudios han revelado que estas estimaciones son a menudo inexactas (quizá porque los médicos tienden a recordar pacientes recientes o destacables y, en consecuencia, a concederles demasiada importancia a la hora de llevar a cabo sus estimaciones).

booksmedicos.org

Para varias enfermedades infecciosas, como la influenza y la infección por *Staphylococcus aureus* resistente a meticilina, los estudios periódicos y los sistemas de seguimiento establecidos por los Centers for Disease Control and Prevention alertan a los médicos sobre el cambio en la prevalencia. Los extensos bancos de datos clínicos informatizados también proporcionan estimaciones cuantitativas de la probabilidad de enfermedad, según determinadas combinaciones de hallazgos clínicos. A pesar de que las estimaciones de prevalencia resultantes probablemente no serán muy precisas, el uso de estimaciones procedentes de la bibliografía médica está destinado a ser más exacto que únicamente el juicio implícito.

Aumentando la probabilidad preprueba de la enfermedad

En el entorno clínico, existen varias maneras en las que se puede aumentar la probabilidad de una enfermedad antes de usar una prueba diagnóstica. Teniendo en cuenta la relación que existe entre el valor predictivo de una prueba y la prevalencia, resulta obvio que aplicar pruebas diagnósticas a pacientes con una mayor probabilidad de sufrir la enfermedad que se está buscando constituye una ventaja para el médico. De hecho, las pruebas diagnósticas son más útiles cuando la presencia de enfermedad no es ni muy probable ni muy improbable.

Características específicas de la situación clínica

Las características específicas de la situación clínica son claramente una fuerte influencia a la hora de decidir qué pruebas deben solicitarse. Los síntomas, los signos y los factores de riesgo de la enfermedad incrementan o disminuyen la probabilidad de encontrarla. Para algunas enfermedades, las reglas de decisión clínica, compuestas por una historia clínica simple y las exploraciones físicas, producen grupos de pacientes con una prevalencia o incidencia conocidas de la enfermedad, como se muestra en el ejemplo de la TVP. Una mujer con dolor torácico tiene más probabilidades de sufrir una enfermedad coronaria si sufre una angina de pecho típica e hipertensión arterial, y si es fumadora. Debido a ello, es más probable que una prueba de esfuerzo anómala represente una enfermedad coronaria en una mujer con estas características que en personas con dolor torácico inespecífico y sin factores de riesgo coronario.

El valor de aplicar pruebas diagnósticas a personas que presentan una mayor probabilidad de sufrir una enfermedad concreta es intuitivamente

booksmedicos.org

obvio para la mayoría de los médicos. A pesar de todo, debido a la mayor disponibilidad de pruebas diagnósticas, es fácil adoptar un enfoque menos selectivo a la hora de solicitar pruebas. Sin embargo, cuanto menos selectivo sea el enfoque, probablemente menor será la prevalencia de la enfermedad y menor será el valor predictivo positivo de la prueba.

Ejemplo

La probabilidad de enfermedad arterial coronaria (EAC), de acuerdo con los resultados de una prueba de ECG con ejercicio (esfuerzo), varía en función de la probabilidad preprueba de EAC, que difiere de acuerdo con la edad, el sexo, los síntomas y los factores de riesgo. La figura 8-8 muestra las probabilidades posprueba tras las pruebas de esfuerzo entre los hombres con diferentes edades, síntomas o factores de riesgo (p. ej., diferentes probabilidades preprueba de EAC). Las curvas elípticas muestran las probabilidades de EAC resultantes después de una prueba con resultados positivos o negativos. Un hombre asintomático de 45 años tiene una probabilidad preprueba de EAC muy baja, que se eleva muy poco (< 10 %) tras una prueba positiva y disminuye casi a cero después de una prueba negativa. En el otro extremo, en un hombre de 55 años con angina típica que tiene una probabilidad preprueba de EAC del 93 %, una prueba positiva eleva la probabilidad de EAC casi al 100 %, mientras que una prueba negativa reduce la probabilidad posprueba a un 75 %, escasamente tranquilizador, ya sea para el paciente o el médico; es probable que sea necesario solicitar más pruebas con independencia del resultado de la prueba en este paciente, haciendo inservible la prueba. La prueba de esfuerzo es más útil en un hombre de 45 años con dolor torácico atípico que tiene una probabilidad preprueba de EAC del 51 %. Una prueba positiva eleva la probabilidad posprueba de EAC a un 75 %, haciendo necesaria una prueba más invasiva y definitiva, mientras que una prueba negativa disminuye la probabilidad de EAC a un 10 %.

booksmedicos.org

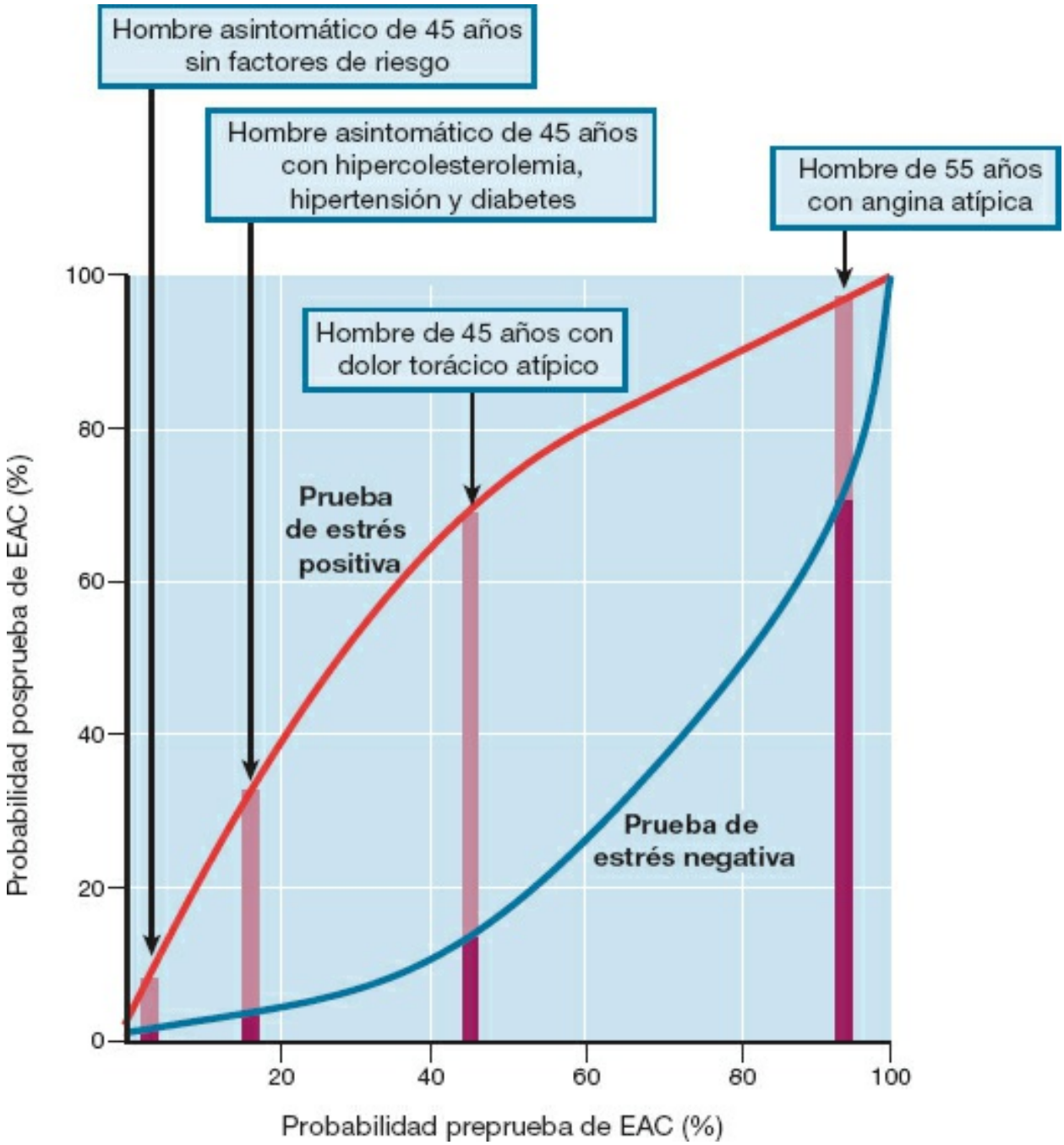


Figura 8-8 ■ Probabilidades posprueba de enfermedad arterial coronaria entre los hombres con diferentes probabilidades preprueba, sometidos a la prueba de ECG con ejercicio (estrés). La parte superior de las barras rosa claro indica las probabilidades posprueba de enfermedad arterial coronaria (EAC) después de las pruebas positivas y la parte superior de las barras rojas indica las probabilidades posprueba de EAC después de las pruebas negativas. (Redibujado con autorización de Patterson RE, Horowitz SF. Importance of epidemiology and biostatistics in deciding clinical strategies for using diagnostic tests: a simplified approach using examples from coronary artery disease. J Am Coll Cardiol 1989;13:1653-1665.)

Debido al efecto de la prevalencia, los médicos deben interpretar de

booksmedicos.org

forma diferente resultados similares de las pruebas en distintas situaciones clínicas. Un resultado positivo habitualmente inducirá a error si se utiliza para buscar una enfermedad no sospechada, como ha sucedido a veces con corredores jóvenes y en revisiones médicas de jóvenes ejecutivos sanos. Justo lo contrario es un hombre mayor con angina de pecho típica. En este caso, un resultado negativo de la prueba de estrés es demasiado a menudo falso negativo para excluir la enfermedad. Como es evidente en la figura 8-8, una prueba diagnóstica es más útil en situaciones intermedias, donde la prevalencia (probabilidad preprueba) no es ni muy alta ni muy baja.

Grupos demográficos seleccionados

En un entorno determinado, los médicos pueden incrementar el rendimiento de las pruebas diagnósticas al realizarlas en grupos demográficos que se sabe que tienen alto riesgo de padecer la enfermedad. La probabilidad preprueba de EAC en un hombre de 55 años que se queja de dolor torácico por angina atípica es del 65 %; en una mujer de 35 años con el mismo tipo de dolor es del 12 % (12). Del mismo modo, una prueba de células falciformes obviamente tendrá un valor predictivo positivo más alto entre los afroamericanos que entre los caucásicos de descendencia noruega.

Proceso de referencia

La derivación a centros hospitalarios docentes, clínicas y departamentos de urgencias aumenta la probabilidad de que una enfermedad grave sea la causa subyacente de las quejas de los pacientes. Por lo tanto, puede que en estos ámbitos esté justificado un uso relativamente más intensivo de las pruebas diagnósticas. (La necesidad de un diagnóstico más rápido también promueve el uso rápido de las pruebas diagnósticas). Por otro lado, en la práctica de atención primaria, y particularmente entre los pacientes sin molestias, la probabilidad de encontrar la enfermedad es considerablemente menor, y se deben utilizar las pruebas con más moderación.

Ejemplo

Durante las prácticas en una clínica militar, uno de nosotros atendió a cientos de personas con cefalea, para las que casi nunca se solicitaban pruebas diagnósticas. Nunca se encontró en los afectados causa alguna subyacente grave para dichas cefaleas. (Es poco probable que se pasaran por alto dolencias importantes, debido a que la clínica era prácticamente la única fuente de

booksmedicos.org

asistencia médica para esos pacientes y los soldados permanecían en la comunidad militar durante muchos meses.) Sin embargo, durante la primera semana de vuelta a la residencia médica, se detectó un absceso cerebeloso a un paciente que acudió al servicio de urgencias por una cefalea similar a las tratadas en el ejército.

Debido a que los médicos pueden trabajar en diferentes puntos a lo largo del espectro de la prevalencia en varios momentos de sus prácticas clínicas, deben tener en cuenta que puede ser necesario ajustar la intensidad de sus evaluaciones diagnósticas para adaptarse al entorno específico.

Implicaciones para la interpretación de la bibliografía médica

Las descripciones publicadas sobre pruebas diagnósticas a menudo incluyen, además de la sensibilidad y la especificidad, algunas conclusiones sobre la interpretación de un resultado negativo o positivo (valor predictivo de la prueba). Esto se hace, con bastante acierto, para proporcionar información que resulte directamente útil para los médicos. Sin embargo, los datos de estas publicaciones se recogen a menudo en hospitales universitarios en los que la prevalencia de enfermedades graves es relativamente alta. Debido a ello, las afirmaciones sobre el valor predictivo en la bibliografía médica pueden inducir a error cuando la prueba se aplica en ámbitos menos seleccionados. En ocasiones, los autores comparan el rendimiento de una prueba en un número de pacientes con la enfermedad diagnosticada con un número igual de pacientes sin la enfermedad. Esta es una forma eficaz de describir la sensibilidad y la especificidad. Sin embargo, cualquier valor predictivo positivo que se notifique en este tipo de estudios tiene muy poco significado, puesto que dicho valor se habrá determinado en un grupo de pacientes para los que la prevalencia de la enfermedad habrá sido establecida de forma artificial por los investigadores en un 50 %.

RAZONES DE VEROSIMILITUD

Las **razones de verosimilitud (RV)** constituyen una forma alternativa de describir el rendimiento de una prueba diagnóstica. Resumen el mismo tipo de información que la sensibilidad y la especificidad, y pueden utilizarse para calcular la probabilidad de enfermedad después de una prueba con resultado positivo o negativo (valor predictivo positivo o negativo). La principal ventaja de las RV es que pueden emplearse en numerosos niveles de resultados de las pruebas.

booksmedicos.org

Momios o posibilidades *(odds)*

Puesto que la utilización de las RV depende de las posibilidades, para entenderlos es necesario ante todo distinguir entre momios y probabilidad. Los momios y la probabilidad contienen la misma información, pero se expresan de manera diferente. La **probabilidad**, empleada para expresar la sensibilidad, la especificidad y el valor predictivo, es la proporción de personas que presentan una característica determinada, como un resultado positivo a una prueba. Por otro lado, los **momios** son el cociente entre dos probabilidades, la probabilidad de un episodio dividida por 1 – la probabilidad del episodio. Pueden convertirse la una en la otra mediante fórmulas sencillas:

$$\text{Momios} = \text{Probabilidad de un episodio} \div (1 - \text{Probabilidad de dicho episodio})$$
$$\text{Probabilidad} = \text{Momios} \div (1 + \text{Momios})$$

La mayoría de los lectores ya deberían conocer estos términos, puesto que se utilizan en las conversaciones del día a día. Por ejemplo, se puede decir que los momios de que esta noche ganen los *New England Patriots* son de 4:1, o que tienen una probabilidad de ganar del 80 %.

Definiciones

La RV para un valor determinado de una prueba diagnóstica se define como la probabilidad de dicho resultado en las personas con la enfermedad, dividida por la probabilidad del resultado en las personas sin la enfermedad (v. fig. 8-2). Las RV expresan el número de veces en que es más (o menos) probable hallar un resultado en las personas enfermas, en comparación con las personas que no lo están. Si una prueba produce resultados dicotómicos (tanto positivos como negativos), dos RV describen la capacidad de dicha prueba para distinguir entre personas enfermas y las que no lo están. Una razón de verosimilitud positivo (RV+) significa el cociente entre la proporción de personas enfermas con una prueba positiva (sensibilidad) y la proporción de personas no enfermas con una prueba positiva (1 – especificidad). Cuando el resultado de una prueba es negativo, se calcula una razón de verosimilitud negativa (RV–), lo que significa la proporción de personas enfermas con un resultado negativo (1 – sensibilidad) dividida por la proporción de personas no enfermas con un resultado negativo

booksmedicos.org

(especificidad) (v. fig. 8-2).

En el ejemplo de la TVP (v. fig. 8-3), los datos pueden utilizarse para calcular las RV de TVP en presencia de un resultado positivo o negativo (estudio de dímeros D). La probabilidad de hallar una prueba positiva es 2,5 veces mayor en presencia de TVP que en ausencia de dicha enfermedad. Si el estudio de dímeros D fuera negativo, la RV de este resultado negativo sería de 0,03.

Uso de las razones de verosimilitud

Las RV deben utilizarse con las posibilidades, y no con la probabilidad. Por consiguiente, el primer paso es convertir la probabilidad preprueba (prevalencia) a **momios preprueba,** tal como se ha descrito anteriormente:

$$\text{Momio} = \text{Probabilidad de un episodio} \div (1 - \text{Probabilidad de dicho episodio})$$

Las RV pueden usarse para convertir momio preprueba en **momios posprueba,** mediante la siguiente fórmula:

$$\text{Momio preprueba} \times \text{Razón de verosimilitud} = \text{Momio posprueba}$$

El momio posprueba se puede convertir, a su vez, de nuevo en probabilidad, mediante la siguiente fórmula.

$$\text{Probabilidad} = \text{Momios} \div (1 + \text{Momio})$$

En estas relaciones, el momio preprueba contiene la misma información que la probabilidad previa o preprueba (prevalencia); las RV, la misma información que sensibilidad/especificidad, y el momio posprueba, la misma que el valor predictivo positivo (probabilidad posprueba).

¿Por qué usar las razones de verosimilitud?

¿Por qué es necesario dominar el concepto de RV si resulta mucho más difícil de entender que el de prevalencia, sensibilidad, especificidad y valor predictivo? La ventaja principal de las RV es que permiten ir más allá de la simple y pobre clasificación del resultado de una prueba como normal o anormal, tal como suele suceder cuando se describe la exactitud de una

booksmedicos.org

prueba diagnóstica únicamente en términos de sensibilidad y especificidad en un solo punto de corte. Obviamente, es mucho más probable que exista una enfermedad si existe un resultado extremadamente anómalo que uno cercano a los límites. Con las RV, es posible resumir la información que contiene el resultado de una prueba en diferentes niveles. Se pueden definir RV para cada uno de los posibles valores de todo un intervalo. De esta forma, la información representada por el grado de anormalidad no se descarta a favor de tan sólo su presencia o ausencia.

Al calcular las RV a lo largo de un intervalo de resultados, se subsana una limitación de la sensibilidad y la especificidad. En lugar de referirse a la capacidad de la prueba para identificar a todos los pacientes con un resultado determinado o peor, la sensibilidad se refiere a la capacidad de un resultado *concreto* para identificar a las personas con la enfermedad. El mismo principio es aplicable al cálculo de la especificidad. Por lo tanto, en las pruebas con un intervalo de resultados, las RV pueden notificar información a cada nivel. En general, las pruebas con RV muy alejadas de 1,0 se asocian con pocos falsos positivos y pocos falsos negativos (> 10 para RV+ y < 0,1 para RV–), mientras que aquellos con RV cercanos a 1,0 proporcionan resultados mucho menos precisos (2,1 a 5,0 para RV+ y 0,5 a 0,2 para RV–).

En resumen, las RV pueden albergar la habitual y razonable práctica clínica de conceder más importancia a los resultados extremadamente elevados (o bajos) que a los resultados limítrofes a la hora de calcular la probabilidad (o los momios) de que exista una determinada enfermedad.

Ejemplo

Los derrames pleurales en general se evalúan de forma rutinaria cuando la causa no está clara, para determinar si el derrame es un trasudado (fluidos corporales que atraviesan una membrana con pocas células o restos) o un exudado (con material celular y restos), que puede indicar una enfermedad inflamatoria o maligna y a menudo requiere de una evaluación diagnóstica más profunda. Tradicionalmente, una forma de diferenciarlas era medir la cantidad de proteínas en el líquido pleural y el suero; si el cociente de proteínas en el líquido pleural y el suero era > 0,50, se consideraba que el líquido pleural era un exudado. Sin embargo, este único punto de corte puede ocultar información importante sobre el diagnóstico. Esta posibilidad fue explorada en un estudio con 1 448 pacientes con derrames pleurales que antes se habían catalogado como exudados o trasudados (13). La tabla 8-2 muestra la distribución de los resultados del líquido pleural según el cociente de proteínas entre el líquido pleural y el suero, junto con las RV calculadas. No es sorprendente que, para valores más elevados en el cociente de proteínas entre el líquido pleural y el suero, casi todos los especímenes fueran exudados y las RV fueran altas, mientras que en el otro

booksmedicos.org

extremo, lo contrario era verdadero. En general, los cocientes cercanos al punto de corte tradicional mostraron que la distinción entre los exudados y los trasudados era incierta. Además, utilizando el punto de corte tradicional de 0,50, se presentaron más errores de clasificación por debajo que por encima del punto de corte.

Tabla 8-2

Distribución de los cocientes de proteína en líquido pleural y suero en pacientes con exudados y trasudados, con el cálculo de las razones de verosimilitud

Cociente de proteínas entre el líquido pleural y el suero	Número de pacientes con el resultado de la prueba		Razón de verosimilitud
	Exudados	Trasudados	
>0,70	475	1	168,65
0,66-0,70	150	1	53,26
0,61-0,65	117	6	6,92
0,56-0,60	102	12	3,02
0,50-0,55	70	14	1,78
0,46-0,50	47	34	0,49
0,41-0,45	27	34	0,28
0,36-0,40	13	37	0,12
0,31-0,35	8	44	0,06
≤0,30	19	182	0,04

Reproducido con autorización de Heffner JE, Sahn SA, Brown LK. Multilevel likelihood ratios for identifying exudative pleural effusions. Chest 2002;121:1916-1920.

La RV presenta otras ventajas sobre la sensibilidad y la especificidad a la hora de describir el rendimiento de una prueba. La información aportada por esta se resume en forma de un número que corresponde a cada nivel del resultado de la prueba. Como se deduce de la tabla 8-2, la información obtenida de la prueba se resume en un número que corresponde a cada nivel del resultado de la prueba. Asimismo, las RV son especialmente idóneas para describir las posibilidades globales de enfermedad cuando se utilizan

booksmedicos.org

diferentes pruebas diagnósticas (v. a continuación).

Cálculo de las razones de verosimilitud

La figura 8-9 A muestra dos formas de obtener la probabilidad posprueba: a través del cálculo matemático y con un nomograma. En la figura 8-9 B, los cálculos se realizan a partir del ejemplo de la TVP de la figura 8-3 y muestra que la probabilidad posprueba calculada (22 %) es la misma que el valor predictivo positivo calculado de acuerdo con el nomograma de la figura. Si bien el proceso es conceptualmente sencillo y no hay cálculo alguno que, por separado, entrañe dificultad, el esfuerzo global resulta un poco desalentador. Para hacerlo más fácil, varios programas informáticos están disponibles de manera gratuita en internet con calculadoras de pruebas diagnósticas, que también construyen el nomograma asociado. El nomograma muestra la diferencia entre los momios *(odds)* pre y posprueba, pero es necesario contar con acceso fácil al nomograma.

booksmedicos.org

A) Método matemático

1) Convertir la probabilidad preprueba (prevalencia) a momio preprueba

Momio preprueba = prevalencia/(1 – prevalencia)

2) Para obtener el momio posprueba, multiplicar el momio preprueba por la razón de verosimilitud

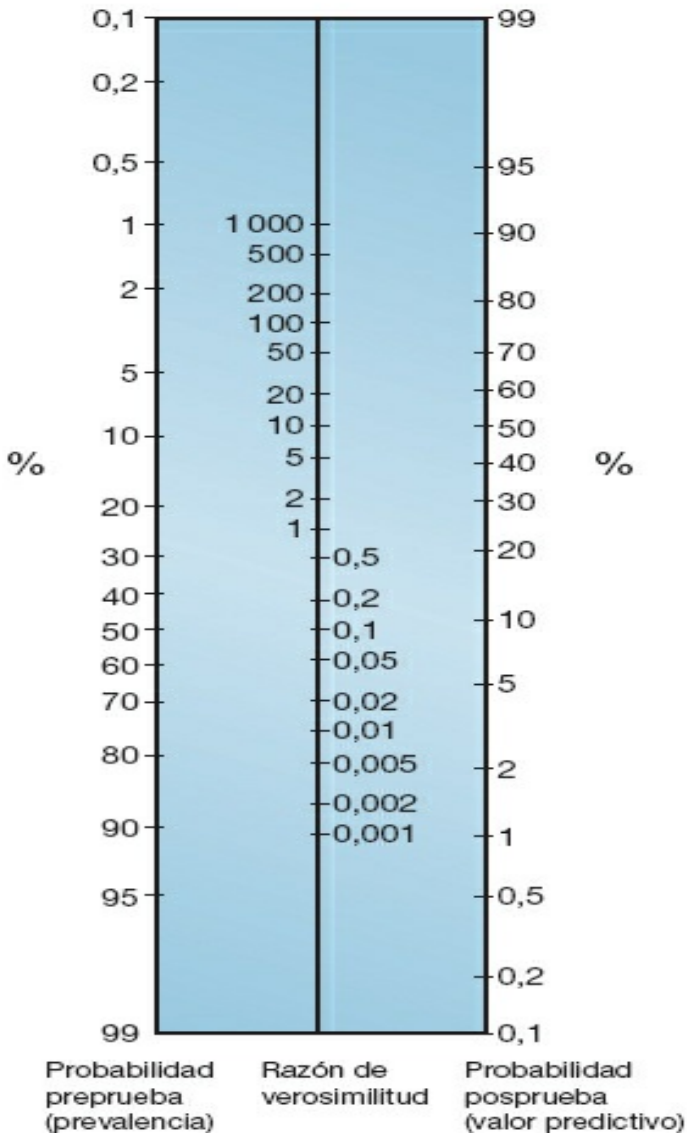
Momio preprueba × razón de verosimilitud = momio posprueba

3) Convertir el momio posprueba a probabilidad posprueba (valor predictivo)

Momio posprueba = momio posprueba/(1 + momio posprueba)

B) Si se utiliza un nomograma de razones de verosimilitud

Trazar una línea recta entre el valor que corresponda a la prevalencia y el que corresponda a la razón de verosimilitud. Al prolongar la línea hacia la probabilidad posprueba se obtendrá dicha probabilidad.



booksmedicos.org

Figura 8-9 **A.** Fórmula y nomograma para determinar la probabilidad posprueba de enfermedad mediante razones de verosimilitud.

booksmedicos.org

A) Método matemático

1) Convertir la probabilidad preprueba (prevalencia) en un momio preprueba

$$0{,}10/(1 - 0{,}10) = 0{,}111$$

2) Multiplicar el momio preprueba por la razón de verosimilitud de una prueba positiva

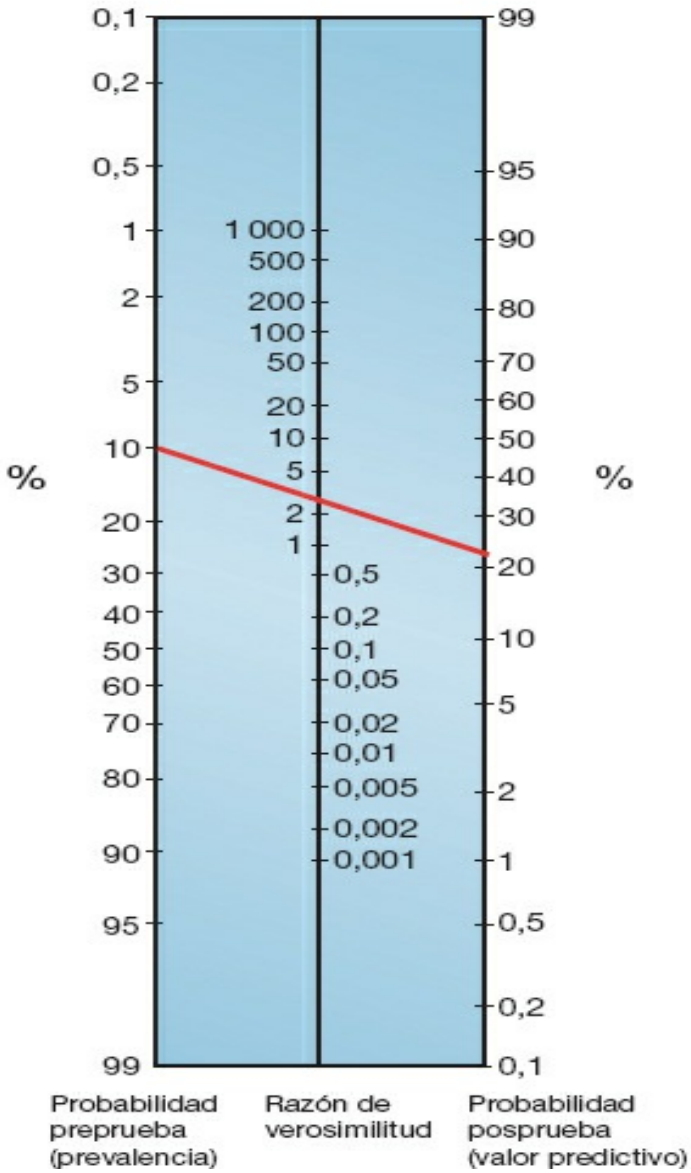
$$0{,}11 \times 2{,}5 = 0{,}278$$

3) Convertir el momio posprueba en probabilidad posprueba (valor predictivo positivo)

$$0{,}278/(1 + 0{,}278) = 0{,}22 = 22\%$$

B) Si se utiliza un nomograma de razones de verosimilitud

La probabilidad preprueba es del 10% y la RV+ es 2,5. Trazar una línea de intersección entre estos dos valores, de forma que cruce la línea de probabilidad posprueba en el valor aproximado del 22%.



booksmedicos.org

Figura 8-9 **B.** Un ejemplo: cálculo de la probabilidad posprueba en un estudio de dímeros D positivo para trombosis venosa profunda (v. fig. 8-3). (Reproducido con autorización de Fagan TJ. Nomogram for Bayes's theorem. N Eng J Med 1975;293:257.)

Tabla 8-3

Una sencilla regla general para determinar el efecto de las razones de verosimilitud sobre la probabilidad de la enfermedad

Razón de verosimilitud	Cambio aproximado en la probabilidad de enfermedad (%)
10	+45
9	+40
8	—
7	—
6	+35
5	+30
4	+25
3	+20
2	+15
1	Sin cambios
0,5	−15
0,4	−20
0,3	−25

booksmedicos.org

0,2	−30
0,1	−45

Adaptado con autorización de McGee S. Simplifying likelihood ratios. J Gen Intern Med 2002;17:646-649.

Estos cálculos muestran un inconveniente de las razones de verosimilitud. Hay que utilizar momios, no probabilidades, y a la mayoría de nosotros nos resulta más dificultoso trabajar con posibilidades que trabajar con probabilidades. Además, la conversión de probabilidad a momios y viceversa precisa cálculos matemáticos o el uso de un nomograma, lo que puede complicar el cálculo de los momios posprueba y de los valores predictivos durante la práctica sistemática de asistencia al paciente.

La tabla 8-3 muestra un enfoque simplificado. Las RV de 2, 5 y 10 incrementan la probabilidad de enfermedad un 15 %, un 30 % y un 45 %, respectivamente, mientras que los cocientes inversos (razones de verosimilitud 0,5, 0,2 y 0,1) reducen la probabilidad de enfermedad de forma parecida, un 15 %, un 30 % y un 45 % (14). El uso clínico de las RV es más sencillo cuando estas tres razones específicas y su efecto (múltiplos de 15) sobre la probabilidad posprueba se recuerdan, especialmente cuando el médico puede calcular la probabilidad de enfermedad del paciente antes de que este se someta a la prueba. Si se aplica este algoritmo al ejemplo de la TVP (fig. 8-3), la probabilidad de la enfermedad con una RV+ de 2,5 sería de un 25 % (prevalencia subyacente del 10 % más un 15 %). Esta estimación es ligeramente superior a la obtenida mediante el cálculo matemático, si bien es lo bastante aproximada como para concluir que un paciente con un estudio positivo de dímeros necesita someterse a alguna otra prueba antes de que se confirme la presencia de TVP.

booksmedicos.org

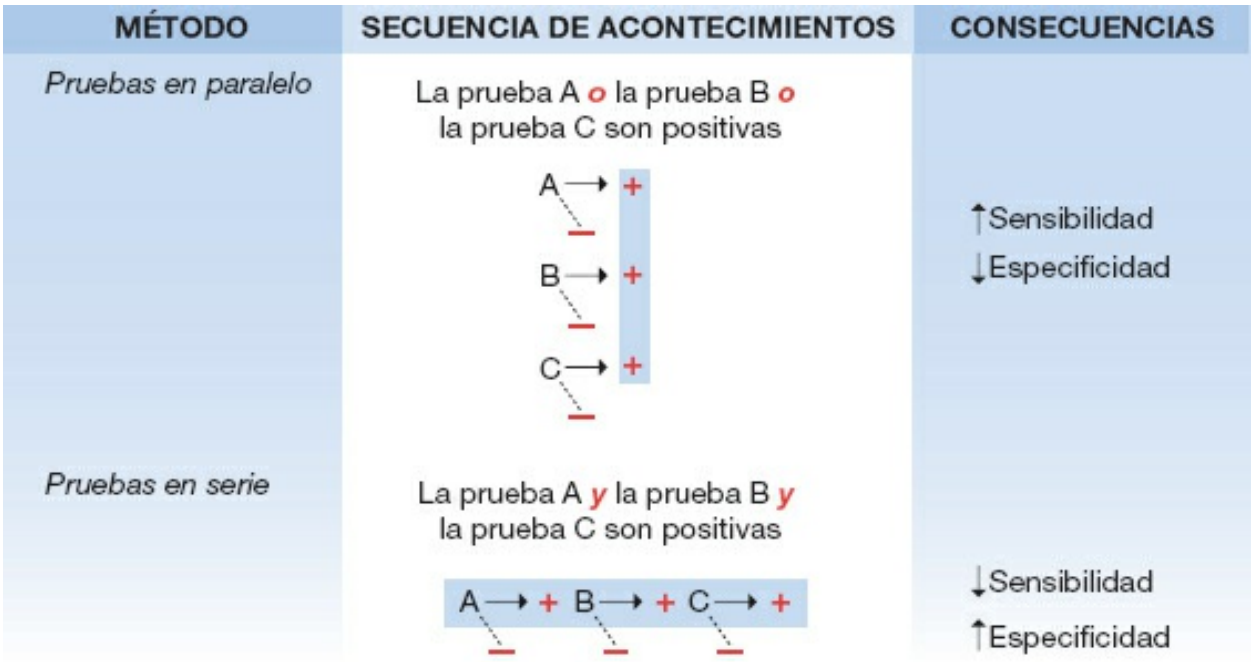


Figura 8-10 **Pruebas en paralelo y en serie.** En las pruebas en paralelo, todas las pruebas se realizan al mismo tiempo. En las pruebas en serie, cada prueba se lleva a cabo sólo cuando el resultado de la prueba anterior ha sido positivo.

MÚLTIPLES PRUEBAS

Debido a que los médicos suelen utilizar pruebas diagnósticas imperfectas con sensibilidades y especificidades inferiores al 100 % y RV intermedias, con frecuencia, una sola prueba da lugar a una probabilidad de enfermedad que no es ni muy alta ni muy baja (p. ej., 10-90 %). Habitualmente, resulta inaceptable interrumpir el proceso diagnóstico en este punto. ¿Acaso quedarían satisfechos un médico o un paciente con la conclusión de que este último todavía tiene un 20 % de posibilidades de sufrir un carcinoma de colon? O que un hombre asintomático de 45 años con múltiples factores de riesgo tiene una probabilidad de un 30 % de enfermedad coronaria después de una prueba de esfuerzo con ECG positiva? Incluso en las enfermedades menos mortales, las pruebas que dan lugar a probabilidades posprueba intermedias requieren más investigación. El médico suele estar obligado a aumentar o disminuir la probabilidad de la enfermedad de forma sustancial, salvo que, naturalmente, las posibilidades diagnósticas sean todas banales, no se pueda hacer nada respecto al resultado o el riesgo de seguir con el proceso sea excesivo. Cuando no concurran estas excepciones, el médico deseará encontrar el modo de confirmar o descartar la enfermedad de una forma más decisiva.

booksmedicos.org

Cuando se realizan varias pruebas diferentes y todas son positivas o negativas, la interpretación no admite dudas, sin embargo, con demasiada frecuencia, algunas pruebas son positivas y otras son negativas. La interpretación se vuelve entonces más complicada. Este apartado aborda los principios que rigen la aplicación e interpretación de múltiples pruebas.

Las diferentes pruebas diagnósticas pueden realizarse básicamente de dos formas (fig. 8-10): **en paralelo** (es decir, todas a la vez), con lo que un resultado positivo se considera un indicio de enfermedad, o pueden realizarse **en serie** (es decir, de forma consecutiva), con lo que la decisión de solicitar la siguiente prueba de la serie se basa en los resultados de la prueba anterior. Para poder efectuar un diagnóstico siguiendo esta modalidad, todas las pruebas deben ser positivas, ya que el proceso diagnóstico se interrumpe cuando se obtiene un resultado negativo.

Pruebas en paralelo

Los médicos suelen solicitar pruebas en paralelo cuando es necesaria una rápida evaluación, como ocurre con pacientes hospitalizados o de urgencias, o en pacientes ambulatorios que no pueden volver con facilidad a su punto de origen, porque han recorrido una larga distancia para someterse a dicha evaluación.

En general, un conjunto de pruebas en paralelo aumenta la sensibilidad y, por lo tanto, el valor predictivo negativo de una determinada prevalencia de enfermedad, por encima de los valores de cada prueba por separado. Por otro lado, la especificidad y el valor predictivo positivo son más bajos que los valores específicos de cada prueba. Es decir, existen menos posibilidades de pasar por alto la enfermedad. (Probablemente, las pruebas en paralelo sean una de las razones que explican por qué los centros de referencia logran diagnosticar las enfermedades que los médicos locales han pasado por alto.) Pero también existen más posibilidades de obtener diagnósticos falsos positivos (de ahí que estos centros también sean propensos al «sobrediagnóstico»). En resumen, las pruebas en paralelo son especialmente útiles cuando el médico se enfrenta con la necesidad de un método diagnóstico muy sensible, y sólo dispone de dos o más pruebas relativamente poco sensibles que miden distintos fenómenos clínicos. Con la utilización de pruebas en paralelo, se obtiene como efecto neto un método diagnóstico más sensible. Sin embargo, en contrapartida, se producen evaluaciones o tratamientos adicionales en pacientes que no sufren la enfermedad.

booksmedicos.org

Tabla 8-4

Características de las pruebas de CA-125 y ecografía transvaginal

Prueba	Sensibilidad[a] (%)	Especificidad[a] (%)	Valor predictivo positivo (%)
CA-125 anormal (≥ 35 U/ml)	55,2	98,7	4,0
Ecografía transvaginal (ETV)	75,8	95,4	1,6
CA-125 anormal o ETV	100,0	94,1	1,7
CA-125 anormal y ETV	31,0	99,9	26,5

[a]Sensibilidad y especificidad estimadas sin cánceres de intervalo (v. texto).
Datos de Buys SS, Partridge E, Greene M, et al. Ovarian cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial: findings from the initial screen of a randomized trial. Am J Obstet Gynecol 2005;193:1630-1639.

Ejemplo

Ninguna de las dos pruebas que se utilizan para diagnosticar el cáncer de ovario, el CA-125 y la ecografía transvaginal, es una prueba sensible cuando se utiliza por separado. Se realizó un estudio donde se realizó a 28 506 mujeres ambas pruebas en un ensayo sobre el tamizaje de cáncer de ovario (15). Si el resultado de alguna de las dos pruebas era anormal (una estrategia de pruebas paralelas), las mujeres eran derivadas para una evaluación más profunda. Se determinaron los valores predictivos positivos para cada prueba individual, así como cuando ambas pruebas fueron anormales (una estrategia de pruebas seriadas). Los autores no calcularon las sensibilidades o especificidades de las pruebas; para hacerlo tendrían que haber incluido cualquier caso de cáncer de ovario que se presentara en las mujeres en el intervalo entre la primera y la siguiente ronda de tamizaje. Sin embargo, casi es posible estimar los valores de la sensibilidad y la especificidad de la información en el documento. La tabla 8-4 muestra los valores predictivos positivos además de las sensibilidades y especificidades estimadas de las pruebas. El uso de las dos pruebas en paralelo elevó la sensibilidad estimada hasta casi el 100 %, pero el valor predictivo positivo fue menor (1,7 %) que al utilizar las pruebas de forma individual. Debido a que la evaluación de seguimiento de las pruebas de tamizaje anormales para el cáncer de ovario a menudo consiste en una cirugía abdominal, son importantes las pruebas con valores predictivos positivos elevados. El valor predictivo positivo bajo, obtenido mediante una estrategia de pruebas paralelas, significó que casi todas las mujeres que requirieron una evaluación de seguimiento, muchas de las cuales fueron sometidas a cirugía, no tuvieron cáncer de ovario. Los autores examinaron lo que habría pasado si se hubiera utilizado una estrategia de pruebas seriadas (que requiere que ambas pruebas sean positivas antes de una evaluación posterior); el valor predictivo positivo aumentó a 26,5 %, pero a costa de disminuir la sensibilidad estimada al 31 %, tan baja que se habrían pasado por alto 20 de 29 casos de cáncer utilizando esta estrategia.

Reglas de predicción clínica

Una variante de las pruebas en paralelo surge cuando los médicos utilizan una combinación de varias pruebas, algunas con resultados positivos y otras con resultados negativos, para llegar a un diagnóstico. Habitualmente, empiezan por un interrogatorio y una exploración física; puede que también soliciten pruebas de laboratorio. A continuación utilizan los resultados

booksmedicos.org

combinados de estas tres fuentes de información para efectuar un diagnóstico. Este proceso, una parte larga e implícita de la medicina clínica, se ha analizado de forma sistemática en un número de diagnósticos cada vez mayor. Para algunas afecciones, ciertos elementos del interrogatorio, los hallazgos físicos y los resultados de laboratorio, son especialmente importantes por su poder predictivo a la hora de efectuar un diagnóstico. Las combinaciones resultantes de estas pruebas se denominan **reglas de predicción clínica** o **reglas de toma de decisiones diagnósticas,** que dividen a los pacientes en grupos con prevalencias diferentes.

Ejemplo

La faringitis es una de las quejas más habituales en la práctica de consulta, pero no está claro qué pacientes deben tratarse con antibióticos para los estreptococos del grupo A. La prueba de referencia para el diagnóstico de estreptococos del grupo A es el cultivo faríngeo, pero ya que los resultados no están disponibles inmediatamente (y porque sólo el 10 % de los pacientes adultos con dolor de garganta tiene faringitis estreptocócica), se han buscado otros métodos para disminuir la necesidad de cultivo. Se estudió una regla de predicción clínica (16) con diferentes riesgos de infección estreptocócica de acuerdo con la puntuación (tabla 8-5). Luego, los autores aplicaron seis estrategias diferentes de tratamiento de acuerdo con los resultados de la puntuación y mostraron las ventajas e inconvenientes en términos de sensibilidad, especificidad, uso innecesario de antibióticos, el esfuerzo del personal clínico y el costo. La regla de predicción clínica identificó a los pacientes con una probabilidad bastante baja de estreptococo del grupo A, que no fueron sometidos a estrategias adicionales de diagnóstico o tratamiento, un abordaje que actualmente se ha incorporado en las guías clínicas de varias organizaciones.

Pruebas en serie

Las pruebas en serie aumentan al máximo la especificidad y el valor predictivo positivo, si bien reducen la sensibilidad y el valor predictivo negativo, como lo muestra la tabla 8-4). Una vez realizadas, el médico está más seguro de que los resultados positivos representan enfermedad, pero corre un mayor riesgo de pasar por alto la misma. Las pruebas en serie resultan especialmente útiles cuando ninguna de las pruebas individuales de que dispone el médico ofrece una gran especificidad.

Los médicos suelen recurrir a las pruebas en serie en situaciones clínicas que no precisan una rápida evaluación de los pacientes, como en las consultas privadas y en las consultas externas de los hospitales, en los que se sigue a los pacientes ambulatorios a lo largo del tiempo. (En el ámbito de la atención

booksmedicos.org

de urgencias, el tiempo es el enemigo del método de las pruebas seriadas.) También se utilizan pruebas en serie cuando algunas son caras o peligrosas, de forma que se recurre a estas últimas sólo cuando las más sencillas y seguras sugieren la presencia de enfermedad. Por ejemplo, para detectar embarazos con un elevado riesgo de engendrar un bebé con síndrome de Down, se emplean la edad materna, ciertos análisis de sangre y ecografías. A las madres que, según estas pruebas, presentan un alto riesgo se les ofrece la amniocentesis, aunque ambos conllevan el riesgo de pérdida fetal. Las pruebas en serie precisan el servicio de los laboratorios con menor frecuencia que las pruebas en paralelo, ya que una evaluación adicional depende de los resultados de la prueba anterior. Sin embargo, las pruebas en serie requieren un tiempo mayor, puesto que sólo se solicitan pruebas adicionales cuando se dispone de los resultados de las pruebas anteriores. Habitualmente, la prueba menos arriesgada, menos cruenta, más fácil de realizar y más económica es la que debe realizarse en primer lugar. Si estos factores no desempeñan un papel importante, lo más eficaz suele ser realizar la prueba que ofrezca una mayor especificidad, puesto que implicará que un número menor de pacientes se sometan a pruebas adicionales.

Tabla 8-5

Ejemplo de una regla de toma de decisiones diagnósticas: predictores de faringitis por estreptococo del grupo A. Puntuación modificada de Centor

Criterios	**Puntuación**
Temperatura > 38 °C	1
Ausencia de tos	1
Ganglios cervicales anteriores inflamados y sensibles	1
Inflamación o exudado amigdalino	1
Edad 3-14 años	1 1

booksmedicos.org

15-44 años ≥ 45 años	0 –1
Puntuación	**Probabilidad de un cultivo para EGA (%)**
≤0	1-2,5
1	5-10
2	11-17
3	28-35
≥4	51-53

EGA, estreptococo del grupo A.
Adaptado con autorización de McIsaac WJ, Kellner JD, Aufricht P, et al. Empirical validation of guidelines for the management of pharyngitis in children and adults. JAMA 2004;291:1587-1595.

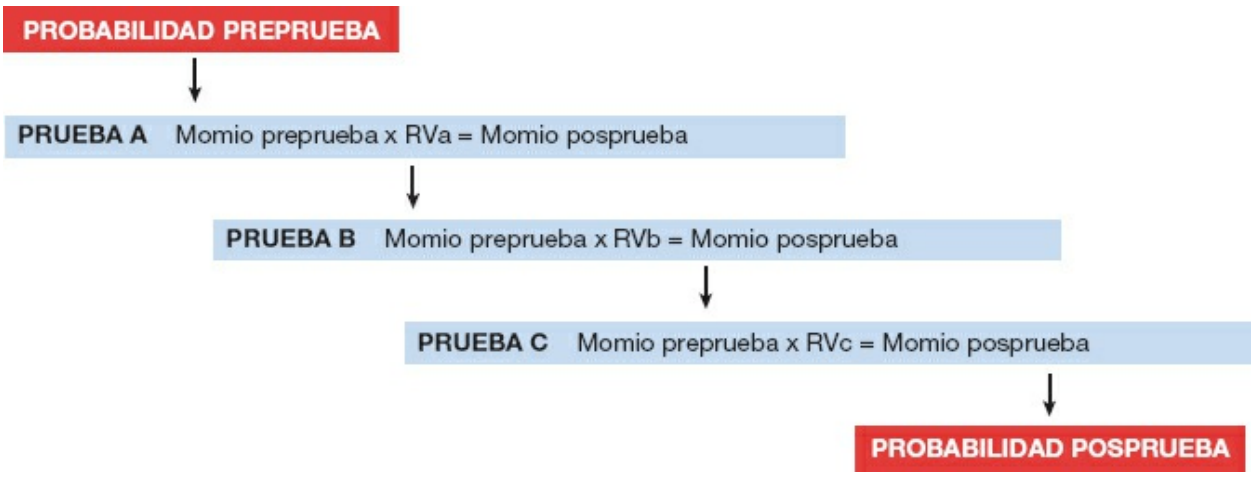


Figura 8-11 ■ Uso de razones de verosimilitud (RV) en pruebas en serie. Conforme se completa una prueba, su momio posprueba se convierte en el momio preprueba de la prueba siguiente.

Razones de verosimilitud en serie

Cuando se utilizan pruebas en serie, puede calcularse una probabilidad global a partir de la RV del resultado de cada prueba, como se muestra en la figura 8-11. En primer lugar, se convierte la prevalencia de la enfermedad antes de

booksmedicos.org

la prueba en un momio preprueba. Conforme se van realizando las pruebas, el momio posprueba de una de ellas pasa a ser el momio preprueba de la prueba siguiente. Al final, se obtiene una nueva probabilidad de la enfermedad que tiene en cuenta toda la información proporcionada por todas las pruebas de la serie.

Supuesto de la independencia

Cuando se emplean varias pruebas, la exactitud del resultado final dependerá de si la información adicional que proporciona cada una de ellas es, en cierto modo, independiente de la información que ya han proporcionado las pruebas anteriores, de forma que la siguiente prueba no se limite simplemente a duplicar la información conocida. Por ejemplo, en el diagnóstico de la endocarditis, es probable que la fiebre (un signo de inflamación), un nuevo soplo cardíaco (un signo de destrucción valvular) y los nódulos de Osler (un signo de la presencia de émbolos) aporten por separado información útil e independiente. En el ejemplo de la faringitis, los investigadores utilizaron técnicas estadísticas para garantizar que cada una de las pruebas diagnósticas incluidas en la regla de decisión aportaba al diagnóstico información independiente. En la medida en que dos pruebas no están aportando información independiente, tiene menor utilidad el uso de múltiples pruebas. Por ejemplo, si se utilizan dos pruebas en paralelo con sensibilidades del 60 % y el 80 %, y la mejor prueba identifica todos los casos detectados por la prueba sensible, la prueba combinada no puede ser superior al 80 %. Si son completamente independientes entre sí, entonces la sensibilidad de las pruebas paralelas sería del 92 % (80 % + [60 % × 20 %]).

La premisa de independencia subyace detrás de todo enfoque para el uso de múltiples pruebas. Sin embargo, parece poco probable que las pruebas para la mayoría de las enfermedades sean completamente independientes entre sí. Si el supuesto de que las pruebas son completamente independientes es erróneo, el cálculo de la probabilidad de enfermedad procedente de varias pruebas tenderá a sobrestimar el valor de estas.

Preguntas de revisión

Las preguntas 8-1 a 8-10 se basan en el siguiente ejemplo. Seleccione la

booksmedicos.org

mejor respuesta para cada pregunta.

Se llevó a cabo un estudio sobre los síntomas y hallazgos físicos en 247 pacientes evaluados por sinusitis. Se realizó el diagnóstico final de acuerdo con los hallazgos de la radiografía (prueba de referencia) (17). Noventa y cinco pacientes tenían sinusitis y 49 de ellos también tenían dolor facial. Ciento cincuenta y dos no tenían sinusitis y 79 de estos tenían dolor facial.

8-1. ¿Cuál es la sensibilidad de la impresión clínica de los médicos respecto a la amigdalitis en este estudio?

A. 38 %.
B. 48 %.
C. 52 %.
D. 61 %.

8-2. ¿Cuál es la especificidad?

A. 38 %.
B. 48 %.
C. 52 %.
D. 61 %.

8-3. Si el doctor pensaba que el paciente tenía sinusitis debido a que el paciente tenía dolor facial, ¿en qué porcentaje de paciente acertaría?

A. 38 %.
B. 48 %.
C. 52 %.
D. 61 %.

8-4. Si el doctor pensaba que el paciente no tenía sinusitis debido a que no tenía dolor facial, ¿en qué porcentaje de pacientes acertaría?

A. 38 %.
B. 48 %.
C. 52 %.
D. 61 %.

booksmedicos.org

8-5. ¿Qué tan habitual fue la sinusitis en este estudio?

A. 38 %.
B. 48 %.
C. 52 %.
D. 61 %.

8-6. ¿Cuál es la probabilidad posprueba de sinusitis en pacientes con dolor facial en este estudio?

A. 38 %.
B. 48 %.
C. 52 %.
D. 61 %.

Las RV positivas y negativas para el dolor facial fueron de 1,0. Cuando los médicos formularon muchas otras preguntas sobre los síntomas de los pacientes y tuvieron en cuenta los hallazgos de la exploración física, las RV para sus impresiones clínicas globales sobre la posibilidad de los pacientes de tener sinusitis fueron las siguientes: «alta probabilidad», RV 4,7; «probabilidad intermedia», RV 1,4 y «baja probabilidad», RV 0,4.

8-7. ¿Cuál es la probabilidad de sinusitis en los pacientes asignados en el grupo de «alta probabilidad» por los médicos?

A. ~10 %.
B. ~20 %.
C. ~45 %.
D. ~75 %.
E. ~90 %.

8-8. ¿Cuál es la probabilidad de sinusitis en los pacientes asignados en el grupo de «probabilidad intermedia» por los médicos?

A. ~10 %.
B. ~20 %.
C. ~45 %.
D. ~75 %.
E. ~90 %.

booksmedicos.org

8-9. ¿Cuál es la probabilidad de sinusitis en los pacientes asignados en el grupo de «probabilidad baja» por los médicos?

A. ~10 %.
B. ~20 %.
C. ~45 %.
D. ~75 %.
E. ~90 %.

8-10. Dadas las respuestas para las preguntas 8-7 a 8-9, ¿cuál de las siguientes afirmaciones es correcta?

A. La impresión clínica de «probabilidad alta» de sinusitis es más útil en el tratamiento de los pacientes que la «probabilidad baja».
B. La impresión clínica de «probabilidad intermedia» de sinusitis es aproximadamente equivalente a lanzar una moneda al aire.
C. El valor predictivo del dolor facial sería mayor en un entorno clínico donde la prevalencia de sinusitis es del 10 %.

Seleccione la mejor respuesta para las preguntas 8-11 y 8-12.

8-11. ¿Cuál de las siguientes afirmaciones es la más correcta?

A. El uso de las pruebas diagnósticas en paralelo aumenta la especificidad y disminuye el valor predictivo positivo.
B. El uso de las pruebas diagnósticas en serie aumenta la sensibilidad y disminuye el valor predictivo positivo.
C. El uso de las pruebas diagnósticas en paralelo aumenta la sensibilidad y el valor predictivo positivo.
D. El uso de las pruebas diagnósticas en serie aumenta la especificidad y el valor predictivo positivo.

8-12. ¿Cuál de las siguientes afirmaciones es la más correcta?

A. Cuando se utilizan pruebas diagnósticas en paralelo o en serie, cada prueba debe aportar información de manera independiente.
B. Cuando se utilizan las pruebas diagnósticas en serie, la prueba con la sensibilidad más baja debe utilizarse primero.
C. Cuando se utilizan pruebas diagnósticas en paralelo, la prueba con

booksmedicos.org

la especificidad más alta debe utilizarse primero.

Encontrará las respuestas en el apéndice A.

BIBLIOGRAFÍA

1. Chobanian AV, Bakris GL, Black HR, et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: The JNC 7 Report. *JAMA* 2003:289:2560–2571.
2. Gann PH, Hennekens CH, Stampfer MJ. A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. *JAMA* 1995;273:289–294.
3. Wheeler SG, Wipf JE, Staiger TO, et al. Approach to the diagnosis and evaluation of low back pain in adults. *In: Basow DS, ed. UpToDate. Waltham, MA; UpToDate;* 2011.
4. Pickhardt PJ, Choi JR, Hwang I, et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N Engl J Med* 2003;349:2191–2200.
5. Bates SM, Kearon C, Crowther M, et al. A diagnostic strategy involving a quantitative latex D-dimer assay reliably excludes deep venous thrombosis. *Ann Intern Med* 2003;138: 787–794.
6. Maisel AS, Krishnaswamy P, Nowak RM, et al. Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. *N Engl J Med* 2002;347:161–167.
7. Steg PG, Joubin L, McCord J, et al. B-type natriuretic peptide and echocardiographic determination of ejection fraction in the diagnosis of congestive heart failure in patients with acute dyspnea. *Chest* 2005;128:21–29.
8. Dearking AC, Aletti GD, McGree ME, et al. How relevant are ACOG and SGO guidelines for referral of adnexal mass? *Obstet Gynecol* 2007;110:841–848.
9. Bobo JK, Lee NC, Thames SF. Findings from 752,081 clinical breast examinations reported to a national screening program from 1995 through 1998. *J Natl Cancer Inst* 2000;92: 971–976.
10. Bates SM, Grand'Maison A, Johnston M, et al. A latex D-dimer reliably excludes venous thromboembolism. *Arch Intern Med* 2001;161:447–453.
11. Wells PS, Owen C, Doucette S, et al. Does this patient have deep vein thrombosis? *The rational clinical examination. JAMA* 2006;295:199–207.
12. Garber AM, Hlatky MA. Stress testing for the diagnosis of coronary heart disease. *In Basow DS, ed. UpToDate. Waltham, MA; UpToDate;* 2012.
13. Heffner JE, Sahn SA, Brown LK. Multilevel likelihood ratios for identifying exudative pleural effusions. *Chest* 2002;121: 1916–1920.
14. McGee S. Simplifying likelihood ratios. *J Gen Intern Med* 2002; 17:646–649.
15. Buys SS, Partridge E, Greene M, et al. Ovarian cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial: findings from the initial screen of a randomized trial. *Am J Obstet Gynecol* 2005;193:1630–1639.
16. McIsaac WJ, Kellner JD, Aufricht P, et al. Empirical validation of guidelines for the management of pharyngitis in children and adults. *JAMA* 2004;291:1587–1595.
17. Williams JW, Simel DL, Roberts L, et al. Clinical evaluation for sinusitis. *Making the diagnosis by history and physical examination. Ann Intern Med* 1992;117:705–710.

booksmedicos.org

Capítulo 9

Tratamiento

Los tratamientos deberían administrarse «no porque tienen que funcionar, sino porque funcionan»

—L.H. Opie
1980

PALABRAS CLAVE

Hipótesis
Tratamiento
Intervención
Efectividad comparativa
Estudios experimentales
Ensayos clínicos
Ensayos controlados y aleatorizados
Equilibrio
Criterios de inclusión
Criterios de exclusión
Comorbilidad
Grandes ensayos simples
Ensayos clínicos prácticos
Ensayos clínicos pragmáticos
Efecto Hawthorne
Placebo
Efecto placebo
Asignación aleatoria
Aleatorización
Características iniciales

booksmedicos.org

Asignación aleatoria estratificada
Cumplimiento
Adherencia
Período de preinclusión
Entrecruzamiento
Cegamiento
Enmascaramiento
Asignación oculta
Simple ciego
Doble ciego
Ensayo abierto
Resultado compuesto
Calidad de vida relacionada con la salud
Estado de salud
Ensayos de eficacia
Ensayos de efectividad
Análisis por intención a tratar
Ensayos explicativos
Ensayos por protocolo
Ensayos de superioridad
Ensayos de no inferioridad
Margen de inferioridad
Ensayos aleatorizados en grupos *(cluster)*
Ensayos cruzados
Ensayos de N igual a 1
Factor de confusión por indicación
Ensayos clínicos de fase I
Ensayos clínicos de fase II
Ensayos clínicos de fase III
Vigilancia poscomercialización

Tras establecer la naturaleza de la enfermedad del paciente y pronosticar el curso habitual, la siguiente pregunta que debemos formularnos es ¿qué se puede hacer al respecto? ¿Existe algún tratamiento que mejore la evolución de la enfermedad? Este capítulo describe la evidencia utilizada para decidir si un tratamiento bienintencionado es realmente eficaz.

IDEAS Y EVIDENCIA

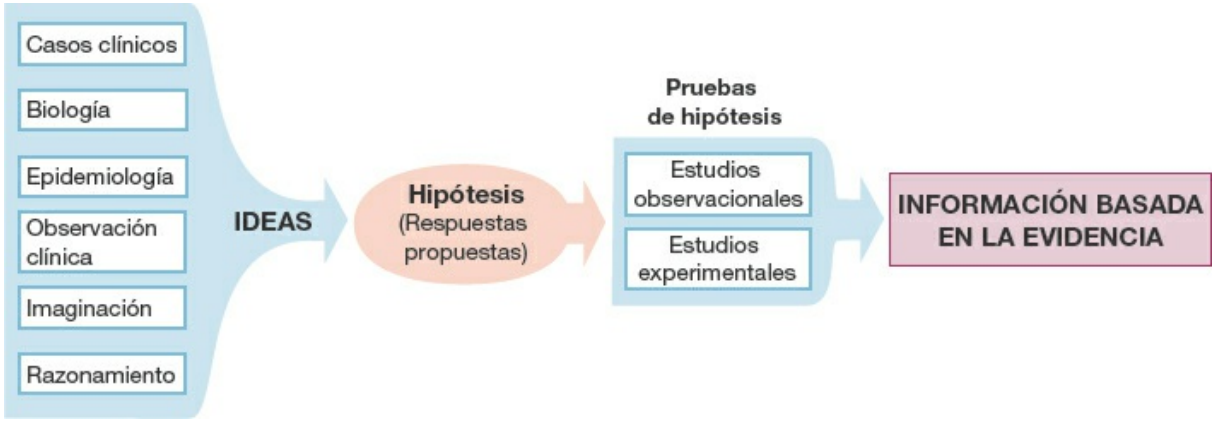
El descubrimiento de nuevos tratamientos requiere buenas fuentes de posibilidades prometedoras y modos rigurosos de establecer que los tratamientos son, de hecho y por término medio, eficaces (fig. 9-1).

Ideas

booksmedicos.org

Las ideas acerca de lo que podría ser un tratamiento útil surgen prácticamente de cualquier actividad dentro de la medicina. Las ideas se denominan **hipótesis** en la medida en que son afirmaciones sobre el mundo natural formuladas para los propósitos de las pruebas empíricas.

Algunas hipótesis terapéuticas las sugieren los mecanismos de la enfermedad a nivel molecular. Los fármacos contra las bacterias resistentes a antibióticos se desarrollan mediante el conocimiento del mecanismo de resistencia. Los análogos hormonales son variaciones de la estructura de las hormonas naturales. Otras hipótesis sobre los tratamientos proceden de observaciones sutiles de los médicos, que se comparten con otros colegas en los casos clínicos. Otras se descubren por accidente: se descubrió que el fármaco minoxidil, que fue creado para la hipertensión, mejora la calvicie de patrón masculino, y tamoxifeno, utilizado para la anticoncepción, prevenía el cáncer de mama en mujeres con riesgo elevado. Las medicinas tradicionales, algunas de las cuales están respaldadas por siglos de experiencia, pueden ser eficaces. La aspirina (ácido acetilsalicílico), la atropina y los digitálicos son ejemplos de sustancias que existen en la naturaleza y que, tras rigurosas pruebas, se han convertido en medicinas ortodoxas. Sin embargo, otras ideas provienen del procedimiento empírico. Se ha demostrado mediante el tamizaje metódico, en modelos de laboratorio, que algunos antineoplásicos tienen una gran cantidad de sustancias activas. Las ideas con respecto al tratamiento, pero con más frecuencia respecto a la prevención, también provienen de estudios epidemiológicos de poblaciones. El estudio Framingham, un estudio de cohorte sobre los factores de riesgo para enfermedades cardiovasculares, fue la base de los ensayos clínicos para disminuir la presión arterial y el colesterol sérico.



booksmedicos.org

Figura 9-1 ■ Ideas y evidencia.

Evaluar ideas

Los efectos de algunos tratamientos son tan rápidos y potentes que su valor es evidente, incluso sin realizar pruebas normales. Los médicos no ponen en duda la efectividad de los antibióticos para la meningitis bacteriana o los diuréticos para el edema. La experiencia clínica es suficiente.

En contraposición, muchas enfermedades, entre ellas la mayor parte de las crónicas, utilizan tratamientos que son mucho menos espectaculares. Los efectos son menores, sobre todo cuando un tratamiento eficaz se prueba contra otro tratamiento eficaz. Además, los resultados tardan más tiempo en presentarse. Por consiguiente, es necesario evaluar formalmente las ideas sobre los tratamientos mediante la investigación clínica porque varias circunstancias, como la coincidencia, las comparaciones sesgadas, los cambios espontáneos en el curso de la enfermedad o una ilusión pueden enmascarar la verdadera relación entre el tratamiento y los resultados.

Cuando se conoce la patogenia de las enfermedades, basado en el trabajo con modelos de laboratorio o estudios fisiológicos en humanos, resulta tentador predecir los efectos en los humanos según este único fundamento. Sin embargo, confiar únicamente en el conocimiento actual de los mecanismos sin comprobar las ideas mediante investigación clínica sólida en humanos sanos puede traer sorpresas desagradables.

Ejemplo

El control del aumento de la glucosa sérica ha sido la piedra angular del tratamiento de los pacientes con diabetes mellitus, en parte para prevenir complicaciones cardiovasculares. La hiperglucemia es la anomalía metabólica más evidente en los pacientes con diabetes. La enfermedad cardiovascular es habitual en los pacientes con diabetes y los estudios observacionales han demostrado la asociación entre la elevación de la glucosa sérica y los episodios cardiovasculares. Para estudiar los efectos del control estricto de la glucosa sérica sobre los episodios cardiovasculares, el ensayo ACCORD aleatorizó a 10 251 pacientes con diabetes mellitus tipo 2 y otros factores de riesgo para enfermedad cardiovascular a recibir el tratamiento intensivo o la atención habitual (1). El control de la glucosa fue sustancialmente mejor en el grupo de tratamiento intensivo pero, sorprendentemente, después de 3,7 años en el ensayo, hubo un 21 % más de muertes en el grupo de pacientes asignados al tratamiento intensivo. También tuvieron más episodios de hipoglucemia y más ganancia de peso. Los resultados fueron similares tras 5 años de seguimiento. El aumento de la mortalidad con el control «estricto», contrario al pensamiento convencional sobre la diabetes (pero consistente

booksmedicos.org

con los resultados de otros ensayos), ha llevado a objetivos menos agresivos para el control de la glucosa sérica y el tratamiento más agresivo de otros factores de riesgo para enfermedad cardiovascular, como el control de la presión arterial, el tabaquismo y la dislipidemia.

Este estudio ilustra cómo los tratamientos de buen juicio, basados en el conocimiento de la enfermedad en ese momento, pueden ser ineficaces cuando se ponen bajo una prueba rigurosa en los humanos. El conocimiento de la patogénesis, desarrollado en modelos de laboratorio, puede ser decepcionante en estudios humanos porque los estudios de laboratorio se llevan a cabo en entornos altamente simplificados. En general excluyen o controlan muchas de las influencias del mundo real sobre la enfermedad, como la variación en la herencia genética, el ambiente físico y social, y las conductas y preferencias individuales.

También se deben poner a prueba la experiencia clínica y la tradición. Por ejemplo, se ha recomendado el reposo en cama para un gran número de trastornos médicos. Generalmente, existe una razón para ello. Por ejemplo, se ha pensado que la cefalea después de una punción lumbar puede ser consecuencia de una pérdida de líquido cefalorraquídeo a través del trayecto de la aguja, lo que causa estiramiento de las meninges. Sin embargo, una revisión de 39 ensayos sobre el reposo en cama para 15 trastornos diferentes encontró que no mejora el resultado en ninguno de los casos. Se observaron peores resultados con el reposo en cama en 17 ensayos, incluyendo no sólo la punción lumbar, sino el dolor lumbar agudo, el parto, la hipertensión durante el embarazo, el infarto agudo de miocardio y la hepatitis aguda infecciosa (2).

Por supuesto, no siempre se da el caso de que las ideas están desacreditadas. El punto principal es que los tratamientos prometedores deben ser probados por la investigación clínica en lugar de ser aceptados para el tratamiento de los pacientes basados sólo en un razonamiento.

ESTUDIOS SOBRE LOS EFECTOS DE LOS TRATAMIENTOS

El **tratamiento** es cualquier intervención cuyo objetivo consiste en mejorar el curso de la enfermedad, una vez establecida. El tratamiento es un caso especial de **intervenciones** en general que se puede aplicar en cualquier momento de la evolución natural de la enfermedad, desde la

booksmedicos.org

prevención de esta hasta el cuidado paliativo al final de la vida. Aunque habitualmente se piensa en las intervenciones como medicamentos, cirugía o radioterapia, estas pueden adoptar cualquier forma, incluyendo la terapia de relajación, la cirugía láser, o los cambios en la organización y financiación de la asistencia médica. A pesar de la naturaleza de una intervención bienintencionada, los principios por los que se considera superior a otras alternativas son los mismos.

La **efectividad comparativa** es un nombre popular para un concepto no tan reciente, la comparación frente a frente de dos o más intervenciones (p. ej., fármacos, dispositivos, pruebas, cirugía o monitorización), que se consideran eficaces y son las opciones actuales de tratamiento. No sólo se compara su efectividad, sino también todos los resultados finales clínicamente importantes de las intervenciones, tanto beneficiosas como dañinas. Los resultados pueden ayudar a los médicos y a los pacientes a comprender todas las consecuencias de elegir una u otra línea de acción cuando ambos se han considerado alternativas razonables.

Estudios observacionales y experimentales de los efectos del tratamiento

Para establecer los efectos de las intervenciones, se utilizan dos métodos generales, los estudios observacionales y los experimentales, que difieren en cuanto a la viabilidad y la rigurosidad científica.

En los estudios obervacionales de intervenciones, los científicos simplemente observan lo que les sucede a los pacientes que por varias razones se exponen o no a una intervención (v. caps. 5-7). Los estudios observacionales de tratamientos constituyen un caso especial de estudios de pronósticos en general, en los que el factor de interés pronóstico es una intervención terapéutica. Lo que se ha comentado sobre los estudios de cohorte también se aplica a los estudios observacionales de tratamientos. La ventaja principal de estos estudios es la viabilidad; el inconveniente fundamental es la posibilidad de que haya diferencias sistemáticas en los grupos de tratamiento, distintas de las del tratamiento en sí, que pueden dar lugar a conclusiones erróneas sobre los efectos del tratamiento.

Los **estudios experimentales** son un tipo especial de estudio de cohortes en el que las condiciones del estudio (es decir, la selección de los grupos de tratamiento, la naturaleza de las intervenciones, el tratamiento durante el seguimiento y la medida de los resultados) las determina el

booksmedicos.org

investigador con el objetivo de establecer comparaciones sin sesgos. Estos estudios se suelen denominar **ensayos clínicos,** y se controlan y organizan mucho más que los estudios de cohortes. Los investigadores llevan a cabo un experimento, análogo a los realizados en el laboratorio. Asumen la responsabilidad (con el permiso de los pacientes) de aislar para el estudio la única contribución de un factor, y mantienen constantes, tanto como sea posible, todos los demás determinantes del resultado.

Los **ensayos aleatorizados y controlados,** en los que el tratamiento se asigna de forma aleatoria, son la pauta por excelencia de los estudios científicos sobre los efectos del tratamiento. A continuación se describen detenidamente, junto con métodos alternativos para estudiar la eficacia de las intervenciones.

ENSAYOS ALEATORIZADOS Y CONTROLADOS

En la figura 9-2 se muestra la estructura de un ensayo aleatorizado y controlado. Todos los elementos son los mismos que para un estudio de cohortes, salvo que el tratamiento se asigna de forma aleatoria, en lugar de ser una elección del médico o el paciente. Las «exposiciones» son tratamientos y los «resultados» son cualquiera de los resultados posibles al final del tratamiento (v. tabla 1-2).

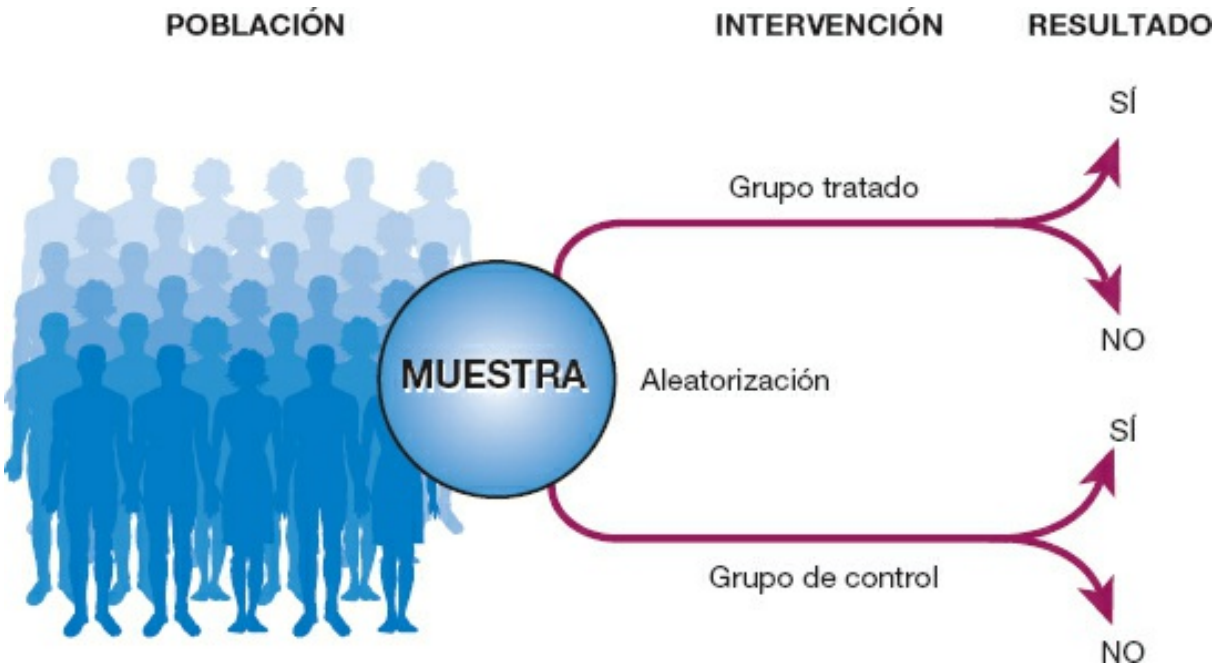


Figura 9-2 Estructura de un ensayo aleatorizado controlado.

booksmedicos.org

Los pacientes que se estudian se seleccionan primero entre una gran cantidad de ellos con la afección de interés. Luego se dividen, mediante aleatorización, en dos grupos (o más) de pronóstico comparable. Uno de los grupos, denominado grupo experimental, se expone a una intervención que se considera mejor que las alternativas actuales. El otro grupo, denominado grupo de control (o grupo de comparación), se trata del mismo modo, pero no expone a sus miembros a la intervención experimental. Los pacientes del grupo de control pueden recibir un placebo, la atención habitual o el mejor tratamiento disponible en el momento. A continuación se registra el curso de la enfermedad en ambos grupos y las diferencias del resultado se atribuyen a la intervención.

El principal motivo de estructurar los ensayos clínicos de este modo es evitar la confusión cuando se comparan los respectivos efectos de dos o más tipos de tratamientos. La validez de los ensayos clínicos depende de cómo se ha creado la igualdad de distribución de todos los determinantes del pronóstico, diferentes del que se está probando en los pacientes tratados y en el grupo de control.

A continuación se describen con detalle los elementos individuales de los ensayos clínicos.

Ética

¿Bajo qué circunstancias es ético asignar un tratamiento de forma aleatoria, en lugar de que lo decidan el paciente y el médico? El principio general, denominado **equilibrio,** es que la aleatorización es ética cuando no hay razones fuertes para creer que cualquiera de los tratamientos asignados al azar es mejor que el otro. Suele pensarse que la intervención experimental *puede* ser mejor que el control, pero esto no se ha establecido de forma concluyente por una investigación sólida. El resultado principal debe ser el beneficio; los tratamientos no pueden ser asignados al azar para descubrir si uno es más dañino que el otro. Por supuesto, como en cualquier otra investigación en humanos, los pacientes deben entender por completo las consecuencias de participar en el estudio, saber que pueden retirarse en cualquier momento sin comprometer su salud, y sentirse libres de expresar su consentimiento para participar. Además, el ensayo debe detenerse siempre que haya evidencia convincente sobre la efectividad, el daño o la inutilidad de continuar.

booksmedicos.org

Muestreo

En los ensayos clínicos, los pacientes deben cumplir unos criterios de inclusión y exclusión rigurosos, que se destinan para incrementar la homogeneidad de los pacientes del estudio, intensificar la validez interna y distinguir fácilmente la «señal» (efecto del tratamiento) del «ruido» (sesgo y azar).

El **criterio de inclusión** habitual consiste en que los pacientes presenten realmente la enfermedad que se estudia. Para estar en el lado seguro, los pacientes del estudio deben cumplir con criterios diagnósticos estrictos. Se puede descartar del proceso a los pacientes con manifestaciones inusuales, leves o dudosas de la enfermedad, con lo que se limita la capacidad de generalización.

De los muchos **criterios de exclusión** que existen, estas son las causas principales de la mayor parte de las pérdidas:

1. Los pacientes con **comorbilidades** u otras enfermedades, aparte de la que se estudia, se excluyen debido a que la atención y la evolución de las otras enfermedades pueden confundir el contraste entre los tratamientos de comparación y experimental, y sus resultados.
2. Se excluye a los pacientes si se considera que su esperanza de vida no es lo suficientemente larga como para presentar los resultados de interés.
3. Los pacientes con contraindicaciones para un tratamiento u otro no pueden participar.
4. Se excluye a los pacientes que rechazan la participación en un ensayo por las razones éticas descritas anteriormente en este capítulo.
5. Los pacientes que no cooperan en las etapas iniciales del ensayo también son excluidos. Esto evita un esfuerzo inútil y la reducción de la validez interna que se produce cuando los pacientes no toman el tratamiento asignado, entran y salen de los grupos de tratamiento o abandonan completamente el ensayo.

Por estos motivos, los pacientes de los ensayos clínicos suelen ser una muestra sesgada muy seleccionada de todos los pacientes con la enfermedad de interés. Dado que se restringe la heterogeneidad, mejora la validez interna del estudio; es decir, hay menos posibilidades de que aparezcan diferencias de resultado que no están relacionadas con el tratamiento en sí. Sin embargo, las exclusiones conllevan la reducción de

booksmedicos.org

la capacidad de generalización: los pacientes en el ensayo no son como la mayoría de los pacientes atendidos en la práctica diaria.

Ejemplo

La figura 9-3 resume cómo se seleccionaron los pacientes para un ensayo aleatorizado y controlado del tratamiento del asma (3). Los investigadores invitaron a participar a 1 410 pacientes con asma de 81 consultas escocesas de medicina general. Sólo 458 pacientes de los invitados, alrededor de un tercio, accedieron a participar y se les podía contactar. Se excluyeron 199, básicamente porque no cumplían los criterios de elegibilidad, y quedaron 259 (un 18 % de los invitados) para ser aleatorizados. Aunque el estudio invitara a pacientes de atención primaria, los que realmente participaron en el ensayo pasaron pruebas de selección muy estrictas, y es poco probable que la mayoría fuera pacientes de ambulatorio.

Debido al elevado nivel de selección de los ensayos, puede ser necesario generalizar los resultados de los ensayos clínicos a los ámbitos de la práctica habitual.

Si no hay suficientes pacientes con la enfermedad de interés, en determinado momento y lugar, para llevar a cabo un ensayo con solidez científica, la muestra puede obtenerse de múltiples sitios con criterios comunes de inclusión y exclusión. Esto se hace principalmente para lograr un tamaño adecuado de la muestra, pero también incrementa la generalización, en la medida que los sitios sean algo diferentes entre sí.

Los **grandes ensayos simples** son una opción para superar el problema de la capacidad de generalización. La inclusión en los ensayos se simplifica, de forma que la mayoría de los pacientes que presentan la enfermedad en estudio son elegibles. Los pacientes que participan tienen que aceptar la asignación aleatoria del tratamiento, pero por lo demás, la asistencia es la misma que reciben habitualmente, sin una gran cantidad de pruebas extras relacionadas con el ensayo. Se realiza el seguimiento para un resultado simple, clínicamente importante, como salir con vida del hospital. Esta propuesta no sólo mejora la capacidad de generalización, sino que también facilita la incorporación de un gran número de participantes con un costo razonable, para poder detectar, dentro de unos límites, las magnitudes del efecto.

Los **ensayos clínicos prácticos** (también conocidos como **ensayos clínicos pragmáticos**) están diseñados para responder preguntas de la vida

booksmedicos.org

real en la atención real de los pacientes, incluyendo el tipo de pacientes e intervenciones que se hallan en los sitios ordinarios de atención para los pacientes.

Ejemplo

Los esguinces de tobillo graves son un problema habitual entre los pacientes que acuden a urgencias. Hoy en día se utilizan varios tipos de tratamiento. El *Collaborative Ankle Support Trial Group* reclutó 584 pacientes con esguinces de tobillo graves en ocho departamentos de urgencias en el Reino Unido, en un ensayo aleatorizado para evaluar cuatro tratamientos habitualmente utilizados: vendaje de compresión tubular y tres tipos de soporte mecánico (4). La calidad de la función del tobillo a los 3 meses fue mejor después de utilizar un yeso por debajo de la rodilla durante 10 días y fue peor cuando se utilizó un vendaje de compresión tubular; en ese momento, el vendaje de compresión tubular se estaba utilizando en el 75 % de los centros en el Reino Unido. Dos formas menos eficaces de soporte mecánico fueron varias veces más costosas que el yeso. Todos los grupos de tratamiento mejoraron con el tiempo y no hubo diferencias en el resultado entre ellos a los 9 meses.

Los ensayos prácticos son diferentes de los típicos de efectividad donde, en un esfuerzo para incrementar la validez interna, se aplican restricciones severas en el reclutamiento, intervención y adherencia, limitando la relevancia de sus resultados para las decisiones de la atención habitual del paciente. Los grandes ensayos simples también pueden ser sobre preguntas prácticas, pero los ensayos prácticos no tienen que ser tan grandes.

Intervención

La intervención en sí puede describirse en función de tres características generales: capacidad de generalización, complejidad y validez.

booksmedicos.org

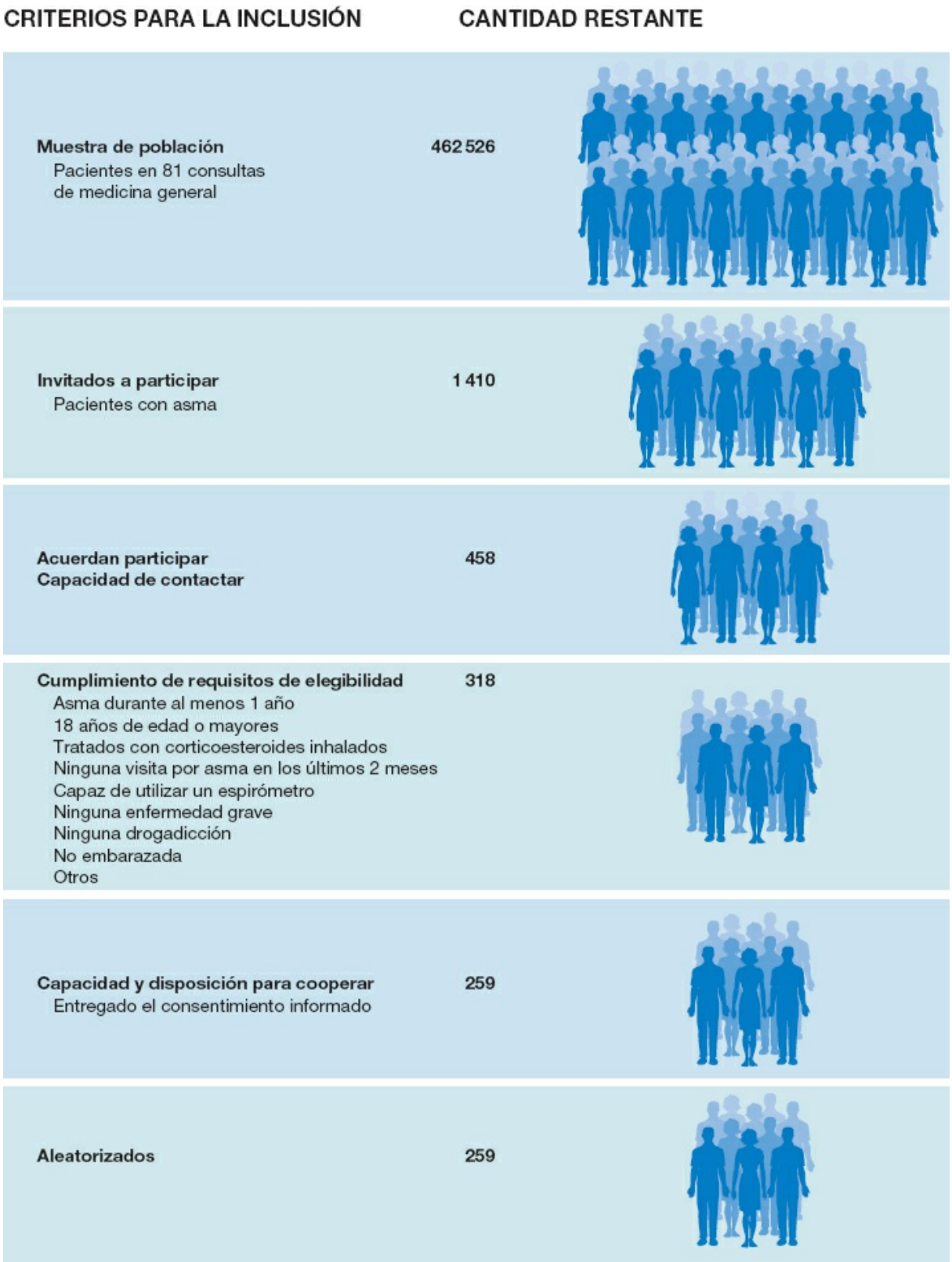


Figura 9-3 ■ Muestreo de pacientes de un ensayo aleatorizado y controlado para el tratamiento del asma. (Datos de Hawkins G, McMahon AD, Twaddle S, et al. Steeping down inhaled corticosteroids in asthma: Randomized controlled trial. BMJ 2003;326:1115-1121.)

booksmedicos.org

En primer lugar, cabe preguntarse si es probable que la intervención se realice en la práctica clínica habitual. En un esfuerzo para estandarizar los tratamientos, de modo que se puedan describir y reproducir fácilmente en otros ámbitos, los investigadores pueden satisfacer a sus compañeros científicos, no a los clínicos, mediante el estudio de tratamientos que no son factibles en la práctica habitual.

En segundo lugar, cabe preguntarse si la intervención refleja una complejidad normal para los planes de tratamiento en el mundo real. Los médicos construyen regularmente planes de tratamiento con varios elementos. Las intervenciones individuales muy específicas conllevan una ciencia ordenada, porque se pueden describir de forma precisa y reproducirse, pero pueden tener efectos poco convincentes. Las intervenciones muy complejas, que suelen ser más eficaces, también pueden evaluarse detalladamente si su esencia se puede comunicar y aplicar en otros marcos. Por ejemplo, un ensayo aleatorizado sobre la prevención de caídas en hospitales de cuidados intensivos estudió los efectos de una escala de evaluación del riesgo de caídas, con intervenciones adaptadas de acuerdo con los riesgos específicos de cada paciente (5).

En tercer lugar, cabe preguntarse si la intervención en cuestión es lo suficientemente diferente de los tratamientos alternativos como para que sea razonable prever que afectará al resultado. Algunas enfermedades se pueden invertir tratando una única causa dominante; un ejemplo es el tratamiento del hipertiroidismo con ablación por radioisótopos o cirugía. Sin embargo, la mayoría de las enfermedades aparecen a partir de una combinación de factores que actúan conjuntamente. No es previsible que las intervenciones que sólo cambian uno de ellos, y sólo una pequeña cantidad, provoquen grandes efectos del tratamiento. No debemos sorprendernos si la conclusión de un ensayo que evalúa estas intervenciones es que el tratamiento nuevo no es eficaz. Por este motivo, los primeros ensayos de un tratamiento nuevo tienden a incluir a los enfermos que tienen más probabilidad de responder al tratamiento, y maximizar las dosis y el cumplimiento.

Grupos de comparación

El valor de una intervención se juzga comparando sus resultados con los de alguna otra actuación alternativa. La pregunta no es si se utiliza un punto

booksmedicos.org

de comparación, sino si este es apropiado para la pregunta de investigación. Los resultados se pueden medir frente a uno o más tipos de grupos de comparación.

- *Sin intervención*. ¿Terminan mejor los pacientes a los que se les ofrece tratamiento experimental que aquellos a los que no se les ofrece nada? Al comparar el tratamiento con la ausencia de este, se miden todos los efectos de la asistencia y de formar parte del estudio, tanto específicos como no específicos.
- *Formar parte de un estudio.* ¿Los pacientes que reciben el tratamiento tienen mejores resultados que otros que acaban de participar en el estudio? En los ensayos clínicos se presta una especial atención a los pacientes. Las personas tienen una tendencia a modificar su conducta cuando son el centro de interés debido al estudio, sin importar la naturaleza específica de la intervención que puedan estar recibiendo. Este fenómeno se denomina **efecto Hawthorne.** Los motivos no están claros, pero hay algunos que parecen probables. Los pacientes tienen ganas de satisfacer a sus médicos y hacerles sentir triunfadores. Además, los pacientes que se prestan voluntarios para los ensayos quieren poner de su parte para ver que se obtienen resultados «buenos».
- *Atención habitual.* ¿Evolucionan mejor los pacientes que reciben tratamiento experimental que los que reciben la atención habitual? sin importar lo que cada médico y los pacientes decidan. Esta es la única pregunta significativa (y ética) si ya se sabe que la atención habitual es eficaz.
- *Tratamiento con placebo*. ¿Evolucionan mejor los pacientes que reciben tratamiento que aquellos a los que se les administra un **placebo,** una intervención que pretende ser idéntica al tratamiento activo (sea por su aspecto físico, color, sabor u olor), pero que no tiene un mecanismo de acción específico conocido? Las pastillas de azúcar y las inyecciones de solución salina son ejemplos de placebo. Se ha demostrado que los placebos, administrados con convicción, alivian síntomas desagradables y graves, como el dolor postoperatorio, las náuseas o el prurito, de alrededor de un tercio de los pacientes, un fenómeno que se denomina **efecto placebo.** Los placebos tienen la ventaja añadida de dificultar a los pacientes del estudio saber el tipo de intervención que han recibido (v. apartado Cegamiento).

booksmedicos.org

- *Otra intervención.* El comparativo puede ser el mejor tratamiento actual. El punto de «efectividad comparativa» es averiguar si el tratamiento nuevo es mejor que el que actualmente se usa.

Los cambios en el resultado relacionados con estos comparativos son acumulativos, como se esquematiza en la figura 9-4.

Asignación del tratamiento

Para estudiar los efectos de una intervención clínica sin otras consecuencias, la mejor forma de asignar pacientes a grupos de tratamiento es mediante la **asignación aleatoria (aleatorización).** Para asignar el tratamiento experimental o el de control a los pacientes se utiliza uno de entre una variedad de procedimientos habituales (lanzar una moneda al aire), a través del cual cada paciente tiene las mismas (o, al menos, que se sepa) posibilidades de ser asignado a cualquiera de los grupos de tratamiento.

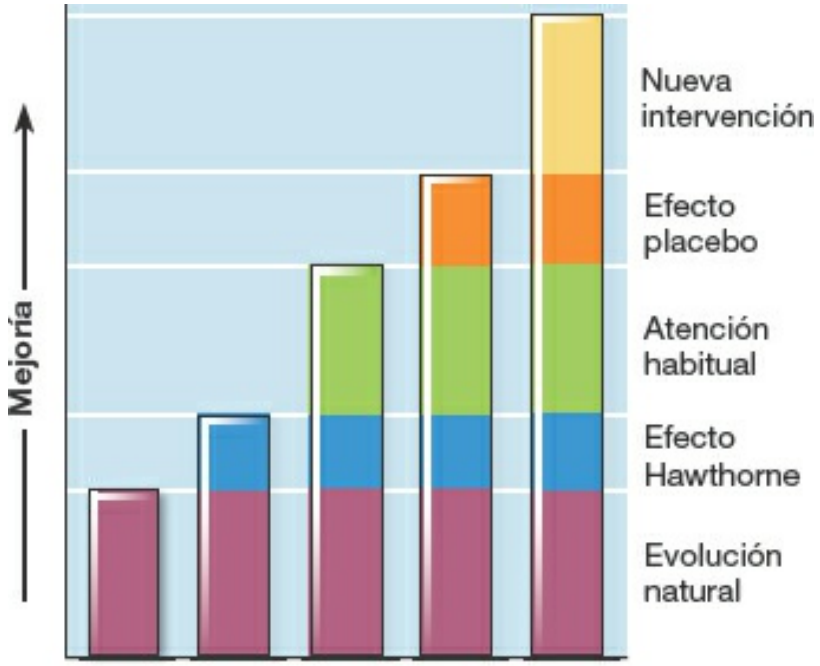


Figura 9-4 Todos los efectos del tratamiento son la suma de la mejoría espontánea (evolución natural), así como las respuestas específicas e inespecíficas.

La asignación aleatoria de los pacientes se prefiere a otros métodos porque sólo la aleatorización tiene la capacidad de crear grupos realmente comparativos. Todos los factores relacionados con el pronóstico, sin

booksmedicos.org

importar si se conocen o no antes de que el estudio se realice o se hayan medido, tienden a distribuirse en partes iguales en los grupos de comparación.

A la larga, en los ensayos de muchos pacientes, la aleatorización suele funcionar como se ha descrito antes. Sin embargo, la asignación aleatoria no garantiza que los grupos sean similares. Pueden surgir diferencias entre los grupos sólo por casualidad, especialmente cuando la cantidad de pacientes aleatorizados es pequeña. Para valorar si se ha dado esta «mala suerte», los autores de ensayos aleatorizados y controlados suelen presentar una tabla en la que se compara la frecuencia de varias características del grupo de tratamiento y del grupo comparativo, en especial las que se sabe que están relacionadas con el resultado. Estas se denominan **características iniciales** porque existen antes de la aleatorización y, por lo tanto, se deben distribuir por igual en los grupos de tratamiento.

Ejemplo

La tabla 9-1 muestra algunas de las características iniciales de un estudio sobre la transfusión sanguínea abundante frente a la transfusión restringida en pacientes de alto riesgo después de una cirugía de cadera. Los 2 016 pacientes reclutados tenían un riesgo mayor debido a factores de riesgo de enfermedad cardiovascular o anemia tras la cirugía. El resultado principal fue la muerte o la incapacidad de caminar en una habitación sin ayuda humana (6). La tabla 9-1 enumera varias características previamente conocidas, a partir de otros estudios o experiencia clínica, que se relacionan con uno o ambos de estos resultados. Cada una de estas características se distribuyó de manera similar en los dos grupos de tratamiento. Estas comparaciones, al menos para las diversas características medidas, fortalecen la idea de que la aleatorización se llevó a cabo de forma correcta y que realmente formó grupos con un riesgo similar de muerte o incapacidad de caminar sin ayuda.

Tabla 9-1

Ejemplo de una tabla de comparación de las características iniciales: ensayo aleatorizado de la transfusión abundante frente a la transfusión restringida en pacientes de alto riesgo tras una cirugía de cadera

		Porcentaje con características para cada grupo

booksmedicos.org

Características	Porcentaje liberal (1 007 pacientes)	Porcentaje restringido (1 009 pacientes)
Edad (promedio)	81,8	81,5
Hombres	24,8	23,7
Cualquier enfermedad cardiovascular	63,3	62,5
Uso de tabaco < 600 mg/día	11,6	11,3
Puntuación de riesgo anestésico	3,0	2,9
Anestesia general	54,0	56,2
Haber vivido en una residencia	10,3	10,9

Datos de Carson JL, Terrin ML, Noveck H, et al. Liberal or restrictive transfusion in high-risk patients after hip surgery. N Engl J Med 2011;365:2453-2462.

Es tranquilizador observar que, de hecho, las características importantes aparecen casi en partes iguales en los grupos que se comparan. Si los grupos son muy diferentes en un ensayo grande, es posible que algo no haya funcionado bien en el proceso de aleatorización. Las diferencias más pequeñas, que se supone que habrá, se pueden controlar durante los análisis de datos (v. cap. 5).

En algunas situaciones, especialmente en ensayos pequeños, para reducir el riesgo de mala suerte, es mejor asegurarse de que, por lo menos, algunas de las características que se sabe que están muy relacionadas con el resultado se dan por igual en los pacientes tratados y en los controles. Los pacientes se distribuyen en grupos (estratos) con niveles similares de un factor pronóstico (p. ej., la edad en la mayoría de las enfermedades crónicas), y se les asigna de forma aleatoria y por separado dentro de cada estrato, un proceso denominado **aleatorización estratificada** (fig. 9-5). Los grupos finales deben ser comparables, como mínimo, respecto a las

booksmedicos.org

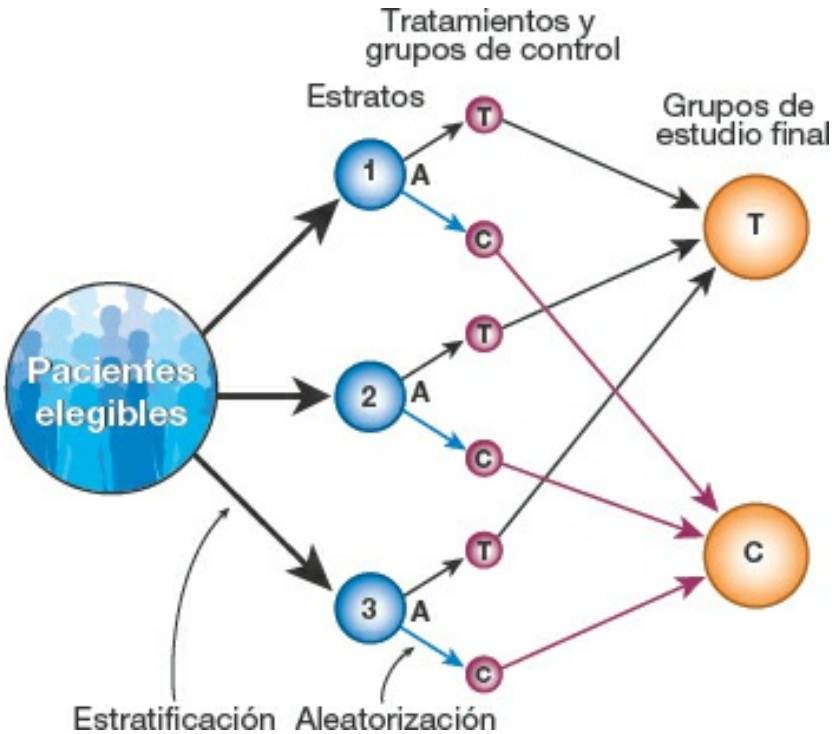
características que se utilizaron para formar los estratos. Algunos investigadores no son partidarios de la aleatorización estratificada, ya que afirman que no es probable que las diferencias que puedan surgir por la "mala suerte" sean muy grandes, y se pueden ajustar matemáticamente tras el control de los datos.

Diferencias tras la aleatorización

No todos los pacientes de un ensayo clínico participan como se planeó en un principio. En algunos se descubre que no sufren la enfermedad que se pensaban que tenían cuando se incorporaron al ensayo. Otros lo abandonan, no toman la medicación, son retirados del estudio debido a los efectos secundarios u otras enfermedades, o de algún modo obtienen otros tratamientos de estudio que no forman parte del trabajo. De esta manera, los grupos de tratamiento que podrían ser comparables justo después de la aleatorización se vuelven menos comparables a medida que pasa el tiempo.

Pacientes sin la enfermedad en estudio

A veces es necesario (tanto en los ensayos clínicos como en la práctica) empezar el tratamiento antes de saber con seguridad si el paciente tiene la enfermedad por la que se le asigna el tratamiento.



booksmedicos.org

Figura 9-5 ■ **Esquema de la asignación aleatoria estratificada.** A, aleatorización; C, grupo de control; T, grupo tratado.

Ejemplo

La hidrocortisona puede mejorar la supervivencia del choque séptico, especialmente en pacientes con una respuesta suprarrenal anormal al choque, según lo medido por un pequeño aumento inapropiado de cortisol plasmático tras la administración de corticotropina. El tratamiento debe empezar antes de tener los resultados de la prueba. Los investigadores en Israel y Europa estudiaron si la hidrocortisona mejoraba la supervivencia a los 28 días en pacientes con choque séptico (7). Los pacientes fueron aleatorizados a recibir hidrocortisona o placebo, y el tratamiento comenzó antes de contar con los resultados de la prueba de corticotropina. De los 499 pacientes con choque séptico reclutados en el ensayo, 233 (47 %) tenían insuficiencia suprarrenal. El análisis principal fue en este subgrupo y no en todos los pacientes aleatorizados, pues se creía que podrían responder a la hidrocortisona. Entre los pacientes con una mala respuesta a la corticotropina, no hubo diferencia en la supervivencia en pacientes tratados con hidrocortisona en comparación con el placebo.

Debido a que la respuesta a la corticotropina fue una característica que existía antes de la aleatorización, y se había asignado al azar, por lo que se conservaron las ventajas de un ensayo clínico controlado y aleatorizado. Sin embargo, no se pudo evitar la ineficacia de reclutar y reunir los datos de los pacientes que no contribuirían a los resultados del estudio. Además, el número de pacientes en el estudio se redujo, haciendo más difícil detectar las diferencias en la supervivencia en caso de existir. No obstante, este tipo de ensayo tiene la gran ventaja de que informa sobre las consecuencias de una decisión que debe tomar un médico antes de tener toda la información relevante, así como la efectividad del tratamiento en el subgrupo de pacientes con mayor probabilidad de responder (v. apartado Intención a tratar y ensayos exploratorios).

Cumplimiento

El **cumplimiento** es la obediencia de las prescripciones médicas por parte del paciente. Puede que sea preferible el término **adherencia** porque tiene una connotación menos sumisa del paciente con respecto al médico. El cumplimiento es otra característica que interviene después de la aleatorización.

Aunque el incumplimiento sugiere un tipo de negligencia intencionada de las recomendaciones, también contribuyen a él otros factores. Los

booksmedicos.org

pacientes pueden malinterpretar los fármacos y las dosis que se han propuesto, se les pueden acabar los medicamentos, pueden confundir varios preparados del mismo fármaco o pueden no tener recursos para pagar los fármacos ni seguro para financiarlos. Si se dan todas estas posibilidades juntas, el incumplimiento puede limitar la utilidad de los tratamientos que han demostrado ser eficaces en condiciones favorables.

En general, el cumplimiento señala un mejor pronóstico, sin considerar el tratamiento. Los pacientes en ensayos aleatorizados que mostraron cumplimiento con el placebo tuvieron mejores resultados que los que no lo hicieron (8).

El cumplimiento es especialmente importante en la asistencia médica extrahospitalaria. En los hospitales actúan muchos factores para controlar la conducta de los pacientes y hacer que estos cumplan con el tratamiento. Los pacientes hospitalizados suelen estar más enfermos y más asustados, se encuentran en un ambiente extraño, dependen de las aptitudes y la atención del personal para todo, incluida su vida. Además, los médicos, el personal de enfermería y los farmacéuticos han desarrollado un sistema bien organizado para garantizar que los pacientes reciban lo que se les prescribe. Como consecuencia, la experiencia clínica y las publicaciones médicas sobre los centros pueden subestimar la importancia del cumplimiento terapéutico fuera del hospital, donde se encuentran la mayoría de los pacientes y médicos, y donde el cumplimiento de las prescripciones es menos frecuente.

En los ensayos clínicos, los pacientes suelen seleccionarse según su cumplimiento. Durante un **período de preinclusión,** en el que se administra un placebo y se supervisa el cumplimiento, se pueden detectar los pacientes que no lo respetan y excluirlos antes de la aleatorización.

Estudios o diseños con grupos cruzados

Los pacientes pueden pasar de un tratamiento asignado al azar a otro durante el seguimiento, un fenómeno denominado **entrecruzamiento** *(cross-over).* Si el intercambio entre los grupos de tratamiento se lleva a cabo a gran escala, puede disminuir las diferencias observadas en el efecto del tratamiento en comparación con lo que podría haberse observado si los grupos originales hubieran permanecido intactos.

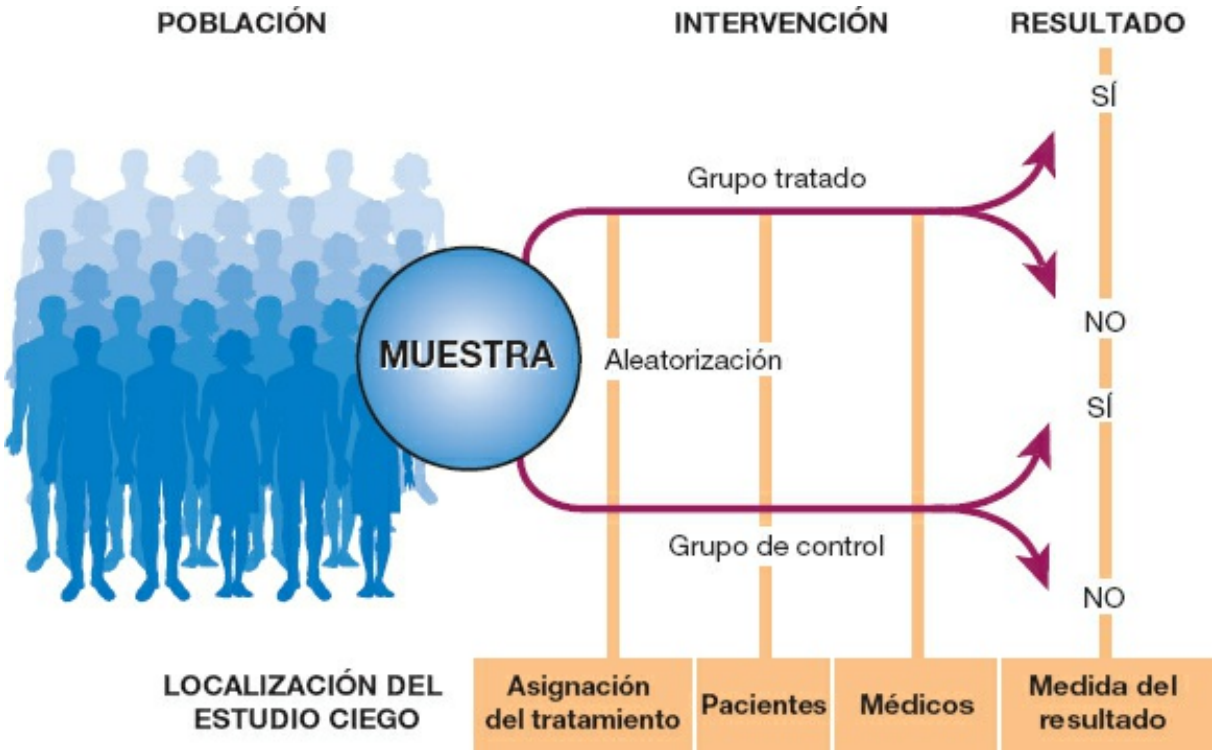
Intervenciones conjuntas

booksmedicos.org

Tras la aleatorización, los pacientes pueden recibir varias intervenciones diferentes a las que se están estudiando. Por ejemplo, en un estudio sobre el tratamiento del asma, pueden recibir no sólo el tratamiento experimental, sino también dosis diferentes de sus medicamentos habituales y hacer esfuerzos mayores para controlar los alergenos en el hogar. Si esto se produce de forma desigual en los dos grupos y afecta a los resultados, pueden aparecer diferencias sistemáticas (sesgos) entre los grupos que no existían cuando estos se formaron.

Cegamiento

Los participantes de un ensayo pueden cambiar de conducta o informar de resultados de manera sistemática (es decir, con sesgos) si saben qué pacientes reciben tal tratamiento. Una forma de reducir al mínimo este efecto es mediante el **cegamiento,** un intento para que los participantes del estudio desconozcan el tratamiento que se ha asignado de forma aleatoria a los pacientes del grupo, para que este conocimiento no les haga actuar de modo diferente y, en consecuencia, disminuya la validez interna del estudio. **Enmascaramiento** es un término más apropiado, pero cegamiento o «a ciegas» es el más utilizado desde hace tiempo.



booksmedicos.org

Figura 9-6 ■ Localizaciones del posible enmascaramiento en los ensayos aleatorizados y controlados.

En un ensayo clínico, la modalidad de cegamiento puede realizarse de cuatro maneras (fig. 9-6). En primer lugar, los encargados de distribuir a los pacientes en grupos de tratamiento no deben saber el tratamiento que se asignará a continuación, para que el hecho de conocerlo no les condicione en la aleatorización. La **asignación oculta** es un término habitual para esta forma de cegamiento; sin él, algunos investigadores podrían caer en la tentación de incorporar tarde pacientes al estudio, o asegurarse de que algunos pacientes reciben el tratamiento que más les conviene. En segundo lugar, los pacientes no deben saber qué tratamiento están tomando, para que no puedan cambiar el cumplimiento ni comunicar síntomas porque conocen la información. En tercer lugar, los médicos que atienden a los pacientes del estudio no deben saber qué tratamiento se ha asignado a cada uno, para que, incluso inconscientemente, no les traten de diferente manera. Por último, si los investigadores que valoran las evoluciones desconocen el tratamiento que ha recibido cada paciente, no se pueden alterar las mediciones que realizan.

A veces, se utilizan los términos **cegamiento simple** (pacientes) y **doble ciego,** pero sus significados son ambiguos. Es mejor describir simplemente lo que se hizo. Cuando el estudio no incluye cegamiento se denomina ensayo *abierto,* o en el caso de ensayos de medicamentos, **ensayo de etiqueta abierta.**

En los estudios sobre fármacos a menudo puede usarse el método de cegamiento en estudios sobre los efectos de fármacos, usando un placebo. Sin embargo, por muchas cuestiones clínicas importantes, como los efectos de la cirugía, la radioterapia, la dieta o la organización de la asistencia médica, es difícil, cuando no imposible, cegar el tratamiento a los pacientes y a los médicos.

Incluso cuando el cegamiento parece posible, suele ser más celebrado que útil. Los efectos fisiológicos como una disminución de la frecuencia del pulso con β-bloqueadores y los trastornos digestivos o la somnolencia con otros fármacos pueden indicar a los pacientes si están tomando el principio activo o el placebo.

Valoración de los resultados

booksmedicos.org

Los ensayos controlados aleatorizados son un tipo especial de estudios de cohortes, y lo que ya se ha comentado sobre las medidas del efecto y los sesgos en los estudios de cohortes (v. cap. 5) también se aplican en ellos, al igual que los peligros de sustituir los resultados intermedios por unos clínicamente importantes (v. cap. 1).

Ejemplo

Un nivel bajo de colesterol HDL (lipoproteína de alta densidad) es un factor de riesgo para enfermedad cardiovascular. La niacina aumenta los niveles de HDL y disminuye el colesterol LDL (lipoproteína de baja densidad) y los triglicéridos (ambos también son factores de riesgo); por lo tanto, la ingesta de niacina podría contribuir a la prevención de enfermedad cardiovascular cuando se toma junto a un fármaco hipolipemiante de la familia de las «estatinas», como simvastatina, que disminuye el colesterol LDL pero no el HDL o los triglicéridos. En un ensayo clínico aleatorizado sobre la adición de niacina a simvastatina, los pacientes que tomaron niacina tuvieron un colesterol HDL más elevado y un LDL y triglicéridos más bajos que los pacientes que sólo tomaron simvastatina (9). Sin embargo, las tasas del resultado primario, una combinación de episodios cardiovasculares (v. siguiente texto) fueron las mismas en los dos grupos de tratamiento: 16,4 % en el grupo de niacina y 16,2 en el grupo de placebo. Los autores concluyeron que «no hubo un mayor beneficio clínico al añadir niacina al tratamiento con estatinas… a pesar de las mejoras significativas en los niveles de colesterol HDL y triglicéridos».

Los ensayos clínicos pueden tener como su resultado principal un **resultado compuesto,** un conjunto de resultados que se relacionan entre sí pero que se tratan como una sola variable de resultado. Por ejemplo, en un estudio sobre la reparación percutánea contra la cirugía abierta para la insuficiencia mitral, el resultado compuesto fue la vida (ausencia de la muerte), la cirugía de la válvula mitral o la insuficiencia mitral grave 12 meses después del tratamiento (10). Existen varias ventajas de este método. Los resultados individuales de la composición pueden estar tan fuertemente relacionados entre sí, biológica y clínicamente, que sería artificial considerarlos por separado. La presencia de uno (como la muerte) puede impedir que el otro (como la regurgitación mitral grave) ocurra. Con más formas de experimentar un episodio de resultado, un estudio es más capaz de detectar los efectos del tratamiento (v. cap. 11). El inconveniente de los resultados compuestos es que pueden obscurecer las diferencias en los efectos de los diferentes resultados individuales. Además, uno de los componentes puede explicar la mayor parte del resultado, dando la

booksmedicos.org

impresión de que la intervención afecta también a las demás. Todos estos inconvenientes se pueden superar simplemente examinando el efecto de cada componente del resultado por separado, y también en conjunto.

Además de los resultados «duros» como la supervivencia, la remisión de la enfermedad y la recuperación de la función, en ocasiones un ensayo mide la **calidad de vida relacionada con la salud** mediante medidas generales y compuestas sobre el **estado de salud.** En la tabla 9-2 se muestra una medición simple de la calidad de vida utilizada por un grupo colaborativo de investigadores del cáncer. Esta «escala de rendimiento» combina los síntomas y la función, como la capacidad de caminar. Otras son mucho más extensas; el *Sickness Impact Profile* contiene más de 100 reactivos y 12 categorías. Se han desarrollado otros específicamente para enfermedades específicas. El problema principal es que el valor de un ensayo clínico se fortalece en la medida en que tales medidas se divulgan junto con las medidas duras, como la muerte y la recurrencia de la enfermedad.

Tabla 9-2

Una medición simple de la calidad de vida. Escala de rendimiento del *Eastern Collaborative Oncology Group*

Estado de rendimiento	Definición
0	Asintomático
1	Sintomático, deambula completamente
2	Sintomático, permanece en cama < 50 % del día
3	Sintomático, permanece en cama > 50 % del día
4	Postrado en cama
5	Muerto

En la tabla 9-2 se resumen las opciones para describir la magnitud del efecto en los ensayos clínicos. Estas opciones son parecidas a los informes

booksmedicos.org

de riesgo y pronóstico, pero están relacionadas con el cambio en el resultado, consecuencia de la intervención.

Tabla 9-3

Resumen de los efectos del tratamiento

Resumen medida[a]	Definición
Reducción del riesgo relativo	$\frac{\text{Tasa de episodios del grupo control} - \text{Tasa de episodios del grupo tratado}}{\text{Tasa de episodios del grupo control}}$
Reducción del riesgo absoluto	Tasa de episodios del grupo control – Tasa de episodios del grupo tratado
Número de pacientes que se necesita tratar *O* número necesario a dañar	$\frac{1}{\text{Tasa de episodios del grupo control} - \text{Tasa de episodios en el grupo tratado}}$

[a]Para datos continuos, en los que hay mediciones en el inicio y tras el tratamiento, las medidas análogas se basan en los valores de la media para los grupos de tratamiento y de control, ya sea después del tratamiento o bien para la diferencia entre los valores basales y los valores tras el tratamiento.

EFICACIA Y EFECTIVIDAD

Los ensayos clínicos pueden describirse en función de los resultados de una intervención en situaciones del mundo real o en situaciones ideales (fig. 9-7).

En primer lugar, hay que preguntarse si el tratamiento puede funcionar en circunstancias *ideales*. Los ensayos que responden a esta pregunta se denominan **ensayos de eficacia.** Los elementos de unas circunstancias ideales son los pacientes que aceptan las intervenciones que se les proponen y siguen las instrucciones concienzudamente, la mejor asistencia posible y la ausencia de efectos extraños por otras enfermedades. La mayoría de los ensayos aleatorizados se diseñan de esta manera.

booksmedicos.org

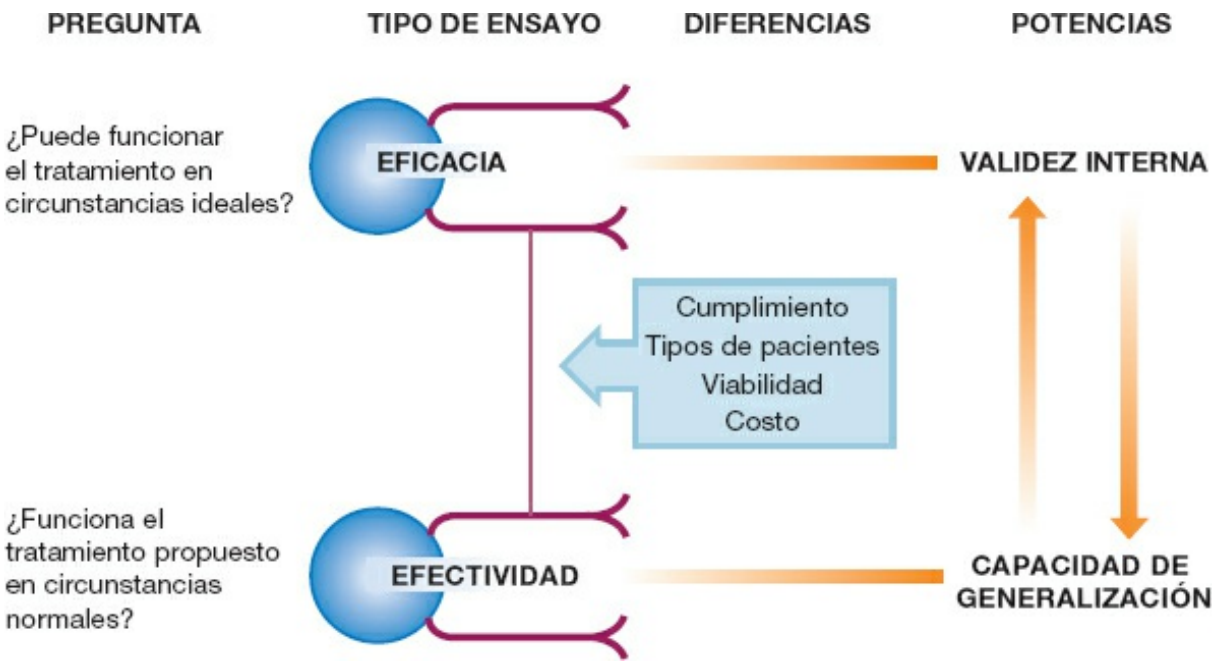


Figura 9-7 ■ Eficacia y efectividad.

En segundo lugar, hay que preguntarse si el tratamiento funciona en circunstancias *normales*. Los ensayos diseñados para responder a esta pregunta se denominan **ensayos de efectividad.** Todos los elementos habituales de la asistencia a pacientes pueden formar parte de los ensayos de efectividad. Es posible que los pacientes no tomen el tratamiento que se les ha asignado. Algunos pueden abandonar el estudio y otros encontrar algún modo de tomar el medicamento que no se les asignó. Los médicos y los centros quizá no sean los mejores. En definitiva, los ensayos de efectividad describen los resultados como la mayor parte de los pacientes los sienten. La diferencia entre la eficacia y la efectividad se ha descrito como la «brecha de la aplicación», la brecha entre la atención ideal y la ordinaria, y es un objetivo para mejorar por sí misma.

Los ensayos de eficacia suelen preceder a los de efectividad. La justificación es que si el tratamiento no es efectivo en las mejores circunstancias, entonces la efectividad será imposible en circunstancias ordinarias. También, si un ensayo de efectividad se hiciera primero y no mostrara ningún efecto, el resultado podría deberse a que el tratamiento en su mejor expresión no es efectivo o porque el tratamiento realmente es efectivo pero no fue administrado.

Intención a tratar y ensayos explicativos

booksmedicos.org

Un tema relacionado es si los resultados de los ensayos aleatorizados y controlados se pueden analizar y presentar según el tratamiento para el que los pacientes fueron aleatorizados o según el que realmente reciben. La presentación correcta de los resultados depende del tema que se plantea (fig. 9-8).

Podemos preguntarnos ¿cuál es la mejor opción terapéutica en el momento en que se debe tomar la decisión? Para responderla, el análisis se realiza en función del grupo al que se asignaron los pacientes (por aleatorización), sin tener en cuenta si realmente se les administró el tratamiento que se suponía debían recibir. Este método de analizar resultados de ensayos se denomina **análisis por intención a tratar,** y una de sus ventajas es que la pregunta corresponde a la que realmente se hacen los médicos; estos ofrecen un tratamiento o no. Además, los grupos de control se aleatorizan desde el principio, de modo que esta comparación tiene toda la validez de un ensayo aleatorizado. El inconveniente es que, cuando muchos pacientes no reciben el tratamiento para el que fueron aleatorizados, las diferencias de efectividad tienden a ser vagas y, por consiguiente, aumentan las posibilidades de observar un pequeño efecto erróneo o ningún efecto estadístico. Si el estudio no muestra diferencia alguna, no estará claro si el tratamiento experimental es en realidad ineficaz o si, sencillamente, no se administró.

booksmedicos.org

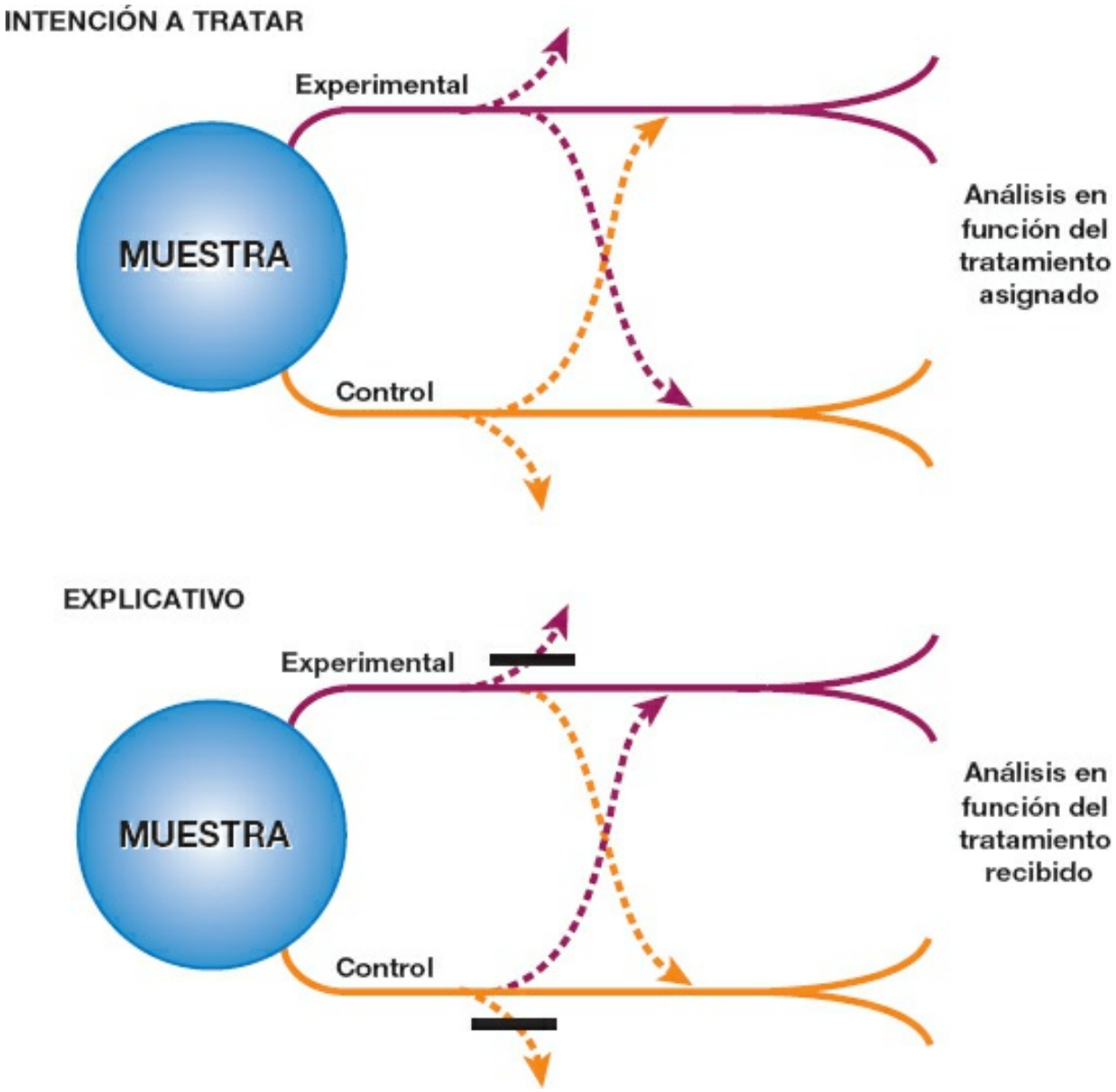


Figura 9-8 ■ Esquema de la asignación de grupo en análisis explicativos y por intención a tratar.

Otra pregunta es si el tratamiento experimental es, en sí mismo, mejor. Para responderla, el análisis se realiza según el tratamiento que cada paciente recibió realmente, sin considerar el tratamiento para el que fueron aleatorizados. Los ensayos que se analizan de este modo se denominan **ensayos explicativos** (también denominados **por protocolo**), porque evalúan si realmente se diferencia el hecho de tomar la medicación o, por el contrario, que sólo se les haya ofrecido. El problema de este enfoque es que salvo que la mayoría de los pacientes reciba el tratamiento que se les ha asignado de forma aleatoria, el estudio ya no representará un ensayo

booksmedicos.org

aleatorizado; es, simplemente, un estudio de cohortes. Hay que tener en cuenta las diferencias entre los grupos, aparte del tratamiento experimental, y se debe utilizar uno o dos métodos, como la restricción, el pareamiento, la estratificación o el reajuste, para lograr que sean comparables, exactamente como se haría para un estudio no experimental.

En general, los análisis por intención a tratar son más relevantes para las preguntas de efectividad, mientras que los análisis explicativos son consistentes con los efectos de los ensayos de eficacia, aunque también importan otros aspectos del ensayo además de cómo son analizados. El análisis primario suele ser por intención a tratar, pero ambos se divulgan. Ambos enfoques son legítimos, y el más adecuado depende de la pregunta que se haya formulado. Estos dos análisis proporcionarán resultados similares, en la medida en que los pacientes de un ensayo sigan el tratamiento que recibieron de forma aleatoria.

SUPERIORIDAD, EQUIVALENCIA Y NO INFERIORIDAD

Hasta ahora se ha discutido sobre los **ensayos de superioridad** que tratan de establecer que un tratamiento es mejor que otro, pero a veces la pregunta más importante es si un tratamiento no es menos eficaz que otro. Un ejemplo clásico es cuando un nuevo medicamento es más seguro, más barato o más fácil de administrar que el establecido y podría, por lo tanto, ser preferible si fuera igual de efectivo. En los **ensayos de no inferioridad,** el propósito es mostrar que es improbable que el nuevo tratamiento sea menos eficaz, al menos de una forma clínicamente importante, que el tratamiento hoy aceptado, que ha demostrado ser más efectivo que el placebo en otros estudios. La pregunta es unidireccional, si un tratamiento no es peor, sin considerar si podría ser mejor.

Desde el punto de vista estadístico, es imposible establecer que un tratamiento no es en absoluto inferior a otro. Sin embargo, un estudio puede descartar un efecto que es menor a una predeterminada «mínima diferencia clínicamente importante», también conocida como **margen de inferioridad,** la diferencia más pequeña en el efecto que aún se considera clínicamente importante. De hecho, este margen tiene en cuenta esta diferencia clínica y la imprecisión estadística del estudio. El siguiente es un ejemplo sobre un ensayo de no inferioridad.

booksmedicos.org

Ejemplo

El pian es una enfermedad infecciosa que afecta a más de 500 000 niños en áreas rurales y pobres de África, Asia y Sudamérica. Es una enfermedad crónica y deformante que afecta sobre todo a la piel, los huesos y el cartílago. La penicilina de acción prolongada por vía parenteral ha sido el fármaco de elección, además de ser barato y tolerarse bien. Sin embargo, existen desventajas de los fármacos inyectables en zonas de escasos recursos, principalmente por las dificultades para organizar el equipo y el personal para administrar las inyecciones y el riesgo que las prácticas inseguras para la inyección puedan provocar la transmisión de patógenos en la sangre, como el VIH o la hepatitis. Para identificar una alternativa efectiva pero más práctica, investigadores de Papúa Nueva Guinea realizaron un ensayo de no inferioridad controlado y aleatorizado sobre la administración oral de una dosis única del antibiótico azitromicina contra una inyección de penicilina (11). La no inferioridad se definió como el límite más bajo del intervalo de confianza al 95 % para la diferencia en las tasas (v. cap. 11), que no era más que un 10 % inferior para la azitromicina que para la penicilina. La tasa de curación fue del 96 % para el grupo de azitromicina y 93 % para el grupo de penicilina, y el intervalo de confianza de la diferencia cumplió con los criterios para que la azitromicina no fuera inferior. Con la evidencia de que un tratamiento más simple es tan efectivo como la inyección de penicilina, la eliminación de la enfermedad, el objetivo en la década de 1960 podría ser realista de nuevo.

Los ensayos de no inferioridad en general requieren un tamaño de la muestra más grande que los ensayos que comparan la superioridad, especialmente si el margen inferior es pequeño o se quieren descartar diferencias pequeñas. Además, cualquier aspecto del ensayo que tiende a minimizar las diferencias entre los grupos de comparación, como los análisis por intención a tratar en los ensayos con el abandono o el entrecruzamiento de muchos pacientes, o cuando las mediciones de los resultados son imprecisas, aumenta de forma artificial la probabilidad de encontrar no inferioridad, sin importar si realmente está presente, es decir, son una prueba débil de no inferioridad.

VARIACIONES EN LOS ENSAYOS ALEATORIZADOS BÁSICOS

En los **ensayos aleatorizados en grupos,** los grupos naturales **(**«***cluster***»)** de pacientes (definidos por los médicos, hospitales, clínicas o comunidades en las que los pacientes están afiliados) son aleatorizados, y los acontecimientos de resultado se expresan en los pacientes de acuerdo con el tratamiento al que fue asignado su grupo. La aleatorización de los

booksmedicos.org

grupos, y no de los individuos en ellos, puede ser preferible por varias razones. Puede que sólo sea más práctico distribuir al azar grupos que personas. Los pacientes dentro de los grupos pueden ser más similares entre sí que los pacientes en otros grupos, y esta fuente de variación, además de los propios tratamientos del estudio, deben tenerse en cuenta en el estudio. Si los pacientes fueran aleatorizados dentro de grupos, ellos o sus médicos podrían saber unos de otros sobre ambos tratamientos y esto podría afectar a su conducta. Por ejemplo, ¿qué tan exitoso sería un médico al tratar aleatoriamente algunos pacientes con infección de las vías urinarias de una forma y a otros pacientes de otra? Del mismo modo, ¿podría un hospital establecer un nuevo plan para prevenir las infecciones de catéteres intravenosos en algunas de sus unidades de cuidados intensivos y no en otras cuando los médicos tratan a los pacientes de ambos sitios a lo largo del tiempo? Por estas razones, la aleatorización es grupos en lugar de los pacientes puede ser el mejor enfoque en algunas circunstancias.

Existen otras variaciones sobre los ensayos aleatorizados controlados habituales («grupo en paralelo»). Los **ensayos cruzados** primero exponen a los pacientes a uno de los dos tratamientos asignados al azar y después al otro. Si se puede suponer que los efectos de la primera exposición ya no están presentes en el momento de la segunda exposición, quizás es porque el tratamiento es de corta duración o porque ha habido un período de descanso entre las exposiciones, cada paciente habrá sido expuesto a cada tratamiento con un orden aleatorio. Esto controla las diferencias entre los pacientes que no están relacionadas con los efectos del tratamiento.

ADAPTANDO LOS RESULTADOS DE LOS ENSAYOS A PACIENTES CONCRETOS

Los ensayos clínicos describen lo que sucede en general. Implica la reunión de muchos pacientes que pueden ser distintos entre sí y los pacientes a quienes se puede generalizar los resultados del ensayo. ¿Cómo se pueden obtener estimaciones sobre el efecto del tratamiento que coincidan más con los pacientes individuales?

Subgrupos

Los pacientes de los ensayos clínicos se pueden dividir en subgrupos, cada

booksmedicos.org

uno con una característica específica (o combinación de características) como la edad, la gravedad de la enfermedad y la comorbilidad, que pueden causar un efecto diferente del tratamiento. Es decir, los datos se examinan para evaluar la modificación del efecto. El número de estos subgrupos está limitado sólo por el número de pacientes en los subgrupos, que ha de ser lo suficientemente elevado como para proporcionar estimaciones bastante estables. Aunque las características utilizadas para definir los subgrupos existieran antes de la aleatorización, los pacientes en cada subgrupo han sido asignados al azar en los grupos de tratamiento. Como consecuencia, los resultados de cada grupo representan, en efecto, un pequeño ensayo dentro de un ensayo. Las características de un paciente dado (p. ej., el paciente puede ser anciano y tener una enfermedad grave pero sin comorbilidad) pueden ser pareadas más específicamente con las de uno de los subgrupos, de lo que podría hacerse con el ensayo como un todo. La efectividad del tratamiento del subgrupo pareado se acercará más a la del paciente individual, y estará limitada básicamente por las conclusiones de los riesgos estadísticos de falsos positivos y falsos negativos, que se describen en el apartado Análisis de subgrupos del capítulo 11.

Efectividad en pacientes individuales

Un tratamiento que en general es efectivo puede no funcionar en un paciente individual. Por lo tanto, los resultados de una investigación clínica válida proporcionan una buena razón para comenzar el tratamiento en un paciente, pero la experiencia con ese paciente es una mejor razón para continuarlo o no. Cuando se trata a un paciente individual, es prudente formular las preguntas siguientes:

- ¿Se sabe (por ensayos aleatorizados controlados) que el tratamiento es eficaz para cualquier paciente?
- ¿Se sabe que el tratamiento es efectivo, en general, en pacientes como el mío?
- ¿El tratamiento está funcionando en mi paciente?
- ¿Los beneficios son dignos de las molestias y los riesgos (de acuerdo con los valores y las preferencias del paciente)?

Al formular estas preguntas y no seguir simplemente sólo los resultados de los ensayos, uno puede protegerse contra la elección del tratamiento sin

booksmedicos.org

fundamentos o la persistencia obstinada a la vista de los malos resultados.

Ensayos de *N* = 1

Los ensayos clínicos rigurosos, que prestan mucha atención a los sesgos y al azar, se pueden realizar con pacientes concretos, de uno en uno. El método, denominado **ensayos de *N* = 1,** es un perfeccionamiento del proceso más informal de ensayo y error. Un paciente recibe un tratamiento u otro, el principio activo o el placebo, en orden aleatorio, ambos durante un período breve. El paciente y el médico desconocen el tratamiento que se administra. Los resultados, como la simple preferencia por un tratamiento o la puntuación de un síntoma, se evalúan tras cada período. Después de muchas repeticiones, los patrones de respuestas se analizan estadísticamente, más de lo que se haría para un ensayo aleatorizado controlado de forma más general. Este método es útil para decidir sobre la atención de pacientes individuales cuando la actividad de la enfermedad es impredecible, la respuesta al tratamiento es rápida y no hay un efecto residual entre un período y otro. Las cefaleas, el asma y la fibromialgia son ejemplos de afecciones para las que se puede emplear este método. Sin embargo, a pesar de su atractivo intelectual, los ensayos *N* = 1 rara vez se llevan a cabo y aún más raramente se publican.

ALTERNATIVAS A LOS ENSAYOS ALEATORIZADOS

Los ensayos aleatorizados controlados son el patrón de referencia para los estudios sobre la efectividad de las intervenciones. Sólo los ensayos aleatorizados grandes pueden eliminar por completo la confusión como una alternativa para explicar los resultados observados.

Limitaciones de los ensayos aleatorizados

Sin embargo, la disponibilidad de varios ensayos aleatorizados controlados bien conducidos no resuelve necesariamente una pregunta. Por ejemplo, después de 21 ensayos aleatorizados controlados durante más de tres décadas (uno se describe en un ejemplo en este capítulo), sigue siendo controvertida la efectividad de los corticoesteroides para el choque séptico. La heterogeneidad de los resultados del ensayo parece estar parcialmente relacionada con las diferencias en la dosis y la duración del medicamento, la proporción de pacientes con insuficiencia suprarrenal relativa y si el

booksmedicos.org

resultado es la reversión del choque o la supervivencia. Es decir, los ensayos eran sobre las mismas preguntas generales pero sobre diferentes preguntas específicas.

Los ensayos clínicos también sufren limitaciones prácticas. Son costosos; en 2011, se estimó que el costo promedio por paciente en los ensayos sobre fármacos era de unos 50 000 dólares, y algunos ensayos más grandes costaron cientos de millones de dólares. La logística puede ser intimidante, sobre todo para mantener métodos similares en todos los sitios de los ensayos multicéntricos y para mantener la integridad del cegamiento de la asignación. La asignación por sí misma sigue siendo un obstáculo, si no del todo para la realización de los ensayos, entonces para la participación total y sin sesgos. Es particularmente difícil convencer a los pacientes para que sean aleatorizados cuando se ha establecido una práctica sin que haya evidencia concluyente sobre su beneficio.

Por estas razones no hay ensayos clínicos disponibles para guiar a los médicos en muchas decisiones importantes sobre el tratamiento, pero aun así se deben tomar las decisiones clínicas. ¿Cuáles son las alternativas de los ensayos aleatorizados controlados? y ¿qué tan fiables son?

Estudios observacionales de intervenciones

A falta de un consenso que favorezca un tipo de tratamiento respecto a otros, se administran varios de acuerdo con las preferencias de cada paciente individual y su médico. Como resultado, en el curso de la asistencia normal al paciente, gran cantidad de pacientes reciben varios tratamientos y continúan manifestando sus efectos. Cuando las experiencias con estos pacientes se recogen y analizan como es debido, se puede complementar la información disponible de los ensayos aleatorizados, sugerir si se requieren nuevos ensayos y proporcionar respuestas cuando los ensayos aún no están disponibles.

Por desgracia, es difícil estar seguro de que los estudios observacionales de tratamiento no muestran factores de confusión. Muchos factores importantes, como la gravedad de la enfermedad, las enfermedades concurrentes, las preferencias locales y la cooperación del paciente determinan las decisiones sobre el tratamiento. Es probable que los pacientes que reciben los distintos tratamientos difieran no sólo en el tratamiento, sino también en otros aspectos.

booksmedicos.org

Ejemplo

El cáncer pulmonar de células no pequeñas es habitual en los ancianos. La quimioterapia paliativa para la enfermedad avanzada alarga la vida; sin embargo, los ancianos reciben quimioterapia con menos frecuencia que los más jóvenes, quizás por la preocupación de que podrían ser más propensos a desarrollar efectos adversos como fiebre e infecciones, lesión nerviosa o trombosis venosa profunda. Para saber si los pacientes ancianos de la comunidad realmente tienen más probabilidades de experimentar efectos adversos, con todo lo demás igual, los investigadores analizaron los datos de un estudio de cohortes sobre la atención y los resultados del cáncer de pulmón (12). Había evidencia consistente de que los pacientes mayores que recibieron quimioterapia se seleccionaron para tener menos probabilidad de presentar efectos adversos: de ellos, menos fueron sometidos a quimioterapia, tuvieron menos efectos adversos antes del tratamiento y recibieron un tratamiento menos agresivo. Aun así, las tasas de efectos adversos después de la quimioterapia fueron de 34-70 % más frecuentes en los pacientes mayores que en los más jóvenes. Este efecto no podía ser atribuido a otras enfermedades más habituales con la edad, que persistieron después de ser ajustadas de acuerdo con la comorbilidad, y es más probable que sean el resultado de la disminución de la función orgánica relacionada con la edad. Una pregunta distinta es si los beneficios del tratamiento superan el aumento en la tasa de efectos adversos.

Especialmente problemático es el **factor de confusión por la indicación** (en ocasiones denominado «causalidad reversa»), que se produce, en el caso de un fármaco, cuando sea lo que sea lo que ha motivado que el médico prescriba un fármaco (las «indicaciones»), ésa es la causa del resultado que se observa, y no sólo el tratamiento por sí mismo. Por ejemplo, puede proponerse a los pacientes un método quirúrgico nuevo porque tiene poco riesgo quirúrgico o no presentan una enfermedad maligna y, por ello, es probable que se puedan beneficiar del procedimiento. En la medida en que se conocen las razones de la elección del tratamiento, se pueden tener en cuenta, al igual que cualquier otro factor de confusión.

Ejemplo

La influenza puede causar un empeoramiento del broncoespasmo en los niños con asma. Se recomienda la vacunación para la influenza, pero a muchos niños con asma no se les ha aplicado, posiblemente por la preocupación de que la vacuna podría causar una exacerbación del asma, que fue sugerido por estudios observacionales. Esta asociación podría haber sido el resultado de la confusión por la indicación mediante el cual los niños con el asma más grave

booksmedicos.org

eran más propensos a ser vacunados y a experimentar más exacerbaciones. Para probar esta posibilidad, los investigadores realizaron un estudio de cohortes basado en la población vacunada para la influenza y las exacerbaciones de asma en los niños de 1 a 6 años de edad en cuatro grandes organizaciones de atención médica durante las tres temporadas de influenza (13). En el análisis crudo (sin ajuste), las tasas de exacerbación del asma fueron de dos a tres veces más altas en los niños vacunados en comparación con los no vacunados. El riesgo era menor, pero seguía siendo elevado, después de ajustar según la edad, el género, la organización de atención médica, la fecha del calendario, la gravedad del asma (basada en el uso de fármacos y las consultas médicas por asma) y las prácticas de atención preventiva. Para lograr un control más preciso de la gravedad del asma, se compararon las tasas en las 2 semanas siguientes a la vacunación con las tasas en los mismos niños en los períodos fuera de este intervalo. En este análisis, cada niño fue su propio control. Los riesgos relativos resultantes mostraron una protección con la vacunación: 0,58, 0,74 y 0,98 en las tres estaciones. Un resumen de los ensayos aleatorizados controlados sugiere que es improbable que la vacuna precipite los ataques de asma (14).

Bases de datos clínicas

A veces se encuentran disponibles bases de datos que incluyen las características iniciales y los resultados de un gran número de pacientes. Los médicos pueden vincular las características de un paciente específico con pacientes similares de la base de datos para determinar cuáles fueron sus resultados. Cuando el uso de una base de datos no es parte de una investigación formal, no se están considerando los factores de confusión y la modificación del efecto, pero la ventaja es que las predicciones provienen de pacientes reales, y no altamente seleccionados como en la mayoría de los ensayos clínicos.

¿Estudios aleatorizados frente a observacionales?

¿Son los estudios observacionales un sustituto fiable para los ensayos aleatorizados controlados? Considerando que los ensayos controlados son el patrón de referencia, la mayoría de los estudios observacionales obtienen la respuesta correcta para la mayoría de las preguntas. Sin embargo, hay excepciones dramáticas. Por ejemplo, los estudios observacionales han demostrado consistentemente que las vitaminas antioxidantes se asocian con un menor riesgo cardiovascular, pero los grandes ensayos aleatorizados controlados no han encontrado este efecto. Por lo tanto, los médicos pueden guiarse por los estudios observacionales sobre los efectos del tratamiento cuando no hay ensayos aleatorizados en los cuales confiar, pero deben mantener un saludable escepticismo.

Los estudios observacionales bien diseñados de las intervenciones

booksmedicos.org

tienen algunas fortalezas que complementan las limitaciones de los ensayos aleatorizados habituales. Consideran los efectos del tratamiento en sí, no sólo de ofrecer tratamiento, que es una pregunta legítima en su propia concepción. En general incluyen a la mayoría de las personas tal como existen en las poblaciones naturales, ya sea en el ámbito clínico o en la comunidad, sin criterios estrictos de inclusión o exclusión. A menudo pueden cumplir con un tiempo de seguimiento más largo que los ensayos, coincidiendo con el tiempo que tarda en desarrollarse la enfermedad y los resultados. Al tener ventaja de los tratamientos y los resultados conforme estos ocurren, los estudios observacionales (sobre todo los estudios de casos y controles y los estudios de cohortes históricas que utilizan registros sanitarios) pueden responder preguntas clínicas más rápidamente que lo que tarda en finalizar un ensayo aleatorizado y, por supuesto, también son menos costosos.

Debido a que el mejor ensayo aleatorizado controlado es el patrón de excelencia, se ha sugerido que los estudios observacionales del tratamiento sean diseñados para parecerse, tanto como sea posible, a un ensayo aleatorizado sobre la misma pregunta (15). Uno podría preguntarse, si el estudio hubiera sido un ensayo aleatorizado, ¿cuáles hubieran sido los criterios de inclusión y exclusión (p. ej., excluir pacientes con contraindicaciones para cualquier intervención), ¿cómo se habría definido exactamente la exposición y cómo se habrían manejado los abandonos y el entrecruzamiento? No se puede esperar que el estudio observacional resultante evite todas las vulnerabilidades, pero al menos sería más fuerte.

FASES DE LOS ENSAYOS CLÍNICOS

En los estudios de fármacos es habitual definir tres fases de los ensayos en el orden en que se efectúan. Los **ensayos de fase I** intentan identificar un intervalo de dosis que se toleren bien y que sean seguras (al menos, para los efectos secundarios graves y muy frecuentes), e incluyen una pequeña cantidad de pacientes (unos 12), sin grupo de control. Los **ensayos de fase II** proporcionan información preliminar respecto a la eficacia del fármaco y a la relación entre la dosis y la eficacia; estos ensayos se pueden comparar, pero cuentan con muy pocos pacientes en los grupos de tratamiento para que se pueda detectar cualquiera de los efectos del tratamiento más largo. Los **ensayos de fase III** son aleatorizados y pueden

booksmedicos.org

proporcionar pruebas definitivas de eficacia e intervalos de efectos secundarios frecuentes. Este tipo de ensayos cuenta con suficientes pacientes, a menudo entre decenas y miles, como para detectar los efectos de los tratamientos clínicamente importantes, y se suelen publicar en revistas biomédicas.

Los ensayos en fase III no son lo suficientemente grandes como para detectar diferencias en la tasa, o incluso la existencia, de efectos secundarios poco frecuentes (v. exposición sobre la potencia estadística en el cap. 11). Por lo tanto, tras lanzar el fármaco al mercado, es necesario hacer el seguimiento de muchos pacientes, un proceso denominado **vigilancia poscomercialización.**

Preguntas de revisión

Lea las siguientes preguntas y seleccione la mejor respuesta.

9-1. Un ensayo aleatorizado controlado compara dos fármacos de uso habitual para el tratamiento del asma. Trescientos pacientes se incluyeron en el ensayo y los criterios de elegibilidad fueron amplios. No se realizó esfuerzo alguno para cegar a los pacientes sobre su grupo de tratamiento después del reclutamiento. Con excepción de los fármacos del estudio, cada médico y su paciente decidieron sobre la atención individualmente. ¿Cuál de las siguientes opciones describe mejor este ensayo?

A. Ensayo clínico práctico.
B. Grandes ensayos simples.
C. Ensayo de eficacia.
D. Ensayos de equivalencia.
E. Ensayo de no inferioridad.

9-2. Un estudio aleatorizado y controlado compara la angioplastia con la fibrinólisis para el tratamiento del infarto agudo de miocardio. Los autores afirman que «el análisis era por intención a tratar». ¿Cuál de los siguientes es una ventaja de este abordaje?

booksmedicos.org

A. Describe los efectos de los tratamientos que los pacientes han recibido en realidad.

B. Es poco probable que se subestime el efecto del tratamiento.

C. No se ve afectado por los pacientes que abandonan el estudio.

D. Describe las consecuencias de proponer tratamientos, tanto si en realidad se siguen tomando como si no.

E. Describe si el tratamiento puede funcionar en circunstancias ideales.

9-3. En un ensayo aleatorizado, los pacientes con meningitis que fueron tratados con corticoesteroides presentaron menores tasas de mortalidad, de hipoacusia y de secuelas neurológicas. ¿Cuál de las siguientes comparaciones es aleatorizada?

A. El subgrupo de los pacientes quienes, en el momento de la aleatorización, estaban severamente afectados por la enfermedad.

B. Los pacientes que sufrieron otros tratamientos frente a los que no las tuvieron.

C. Los pacientes que permanecieron en el ensayo frente a los que lo abandonaron después de la aleatorización.

D. Los pacientes que respondieron al fármaco frente a los que no lo hicieron.

E. Los pacientes que tomaron el fármaco frente a los que no lo tomaron

9-4. Un paciente le pide consejo sobre si debería empezar un programa de ejercicio físico para reducir el riesgo de muerte súbita. Usted busca ensayos aleatorizados y controlados, pero sobre este tema sólo encuentra estudios observacionales. Algunos son estudios de cohortes en los que se comparan las tasas de muerte súbita en pacientes que hacen ejercicio con las tasas de muerte súbita en personas sedentarias; otros son estudios de casos y controles en los que se comparan tipos de ejercicios en personas que han sufrido muerte súbita con controles pareados. Cuál de las siguientes *no* es una ventaja de los estudios observacionales sobre tratamientos, como estos, sobre los ensayos aleatorizados controlados?

A. Los efectos que se notifican son para los pacientes que realmente

booksmedicos.org

han seguido la intervención.

B. Puede que sea posible realizar estos estudios con datos que ya existen, que fue obtenida con otros fines.

C. Capacidad de generalizar a los ámbitos del mundo real más habituales.

D. Los grupos de tratamiento podrían haber tenido un pronóstico similar excepto para el tratamiento en sí.

E. Puede que sea posible realizar estos estudios con datos que ya existen, obtenidos con otros fines.

F. Es más fácil conseguir un gran tamaño de la muestra.

9-5. En un ensayo aleatorizado y controlado de un programa para reducir los problemas de las extremidades inferiores en pacientes con diabetes mellitus, se excluyó a los pacientes si eran menores de 40 años, si les habían diagnosticado la enfermedad antes de cumplir 30 años, si recibieron medicación para la hiperglucemia, si sufrían otras enfermedades o incapacidades graves, o si no respetaban el cumplimiento del tratamiento prescrito durante un período de preinclusión. ¿Cuál de las siguientes es una ventaja de este abordaje?

A. Mejora la capacidad de generalización del estudio.

B. Posibilita la realización de un análisis por intención a tratar.

C. Hace posible un ensayo de efectividad.

D. Evita los sesgos de selección.

E. Mejora la validez interna del estudio.

9-6. ¿Cuál de los siguientes *no* se puede llevar a cabo mediante un análisis por intención a tratar?

A. Una comparación de los efectos de tomar realmente los tratamientos experimentales.

B. Una comparación de los efectos de ofrecer los tratamientos experimentales.

C. Una comparación aleatorizada sobre los efectos del tratamiento.

9-7. Usted está leyendo un informe sobre un ensayo aleatorizado controlado y se pregunta si es probable que la aleatorización estratificada, que se utilizó en el ensayo, pueda mejorar la validez

booksmedicos.org

interna. ¿Para cuál de las siguientes opciones es particularmente útil la aleatorización estratificada?

A. Si el estudio incluye muchos pacientes.
B. Una de las variables iniciales se relaciona fuertemente con el pronóstico.
C. Se espera que muchos pacientes abandonen el estudio.
D. Se planea utilizar un análisis por intención a tratar.

9-8. Un análisis aleatorizado controlado se analiza de acuerdo con el tratamiento que cada paciente realmente recibió. ¿Cuál de los siguientes describe mejor este tipo de análisis?

A. Superioridad.
B. Intención a tratar.
C. Explicativo.
D. Fase I.
E. Abierto.

9-9. En un estudio aleatorizado y controlado se descubre que un β-bloqueador es más eficaz que el placebo para aliviar el miedo escénico. Los pacientes que tomaban el β-bloqueador tendían a presentar una frecuencia de pulso más baja y a sentirse más apáticos, efectos conocidos de estos fármacos. ¿En cuál de los siguientes es posible el cegamiento?

A. Los investigadores que asignaron los pacientes a los grupos de tratamiento.
B. El paciente del ensayo.
C. Los médicos de los pacientes.
D. Los investigadores que evaluaron el resultado.

9-10. ¿Cuál de las siguientes afirmaciones describe mejor el «equilibrio» como la justificación de un ensayo aleatorizado con dos fármacos?

A. Se sabe que los fármacos son igualmente efectivos.
B. Se sabe que uno de los fármacos es más tóxico.
C. Se sabe que ninguno de los fármacos es más efectivo que otro.
D. Aunque un fármaco es más efectivo, el otro fármaco se puede

booksmedicos.org

tomar más fácilmente y tiene menos efectos adversos.

9-11. El antibiótico A es el tratamiento establecido para la neumonía adquirida en la comunidad, pero es costoso y tiene muchos efectos adversos. Se ha desarrollado un nuevo fármaco, el antibiótico B, para la neumonía adquirida en la comunidad que es menos costoso y tiene menos efectos adversos, pero aún no se ha establecido adecuadamente su eficacia en relación con el fármaco A. ¿Cuál de los siguientes sería el mejor tipo de ensayo para evaluar el fármaco B?

A. Superioridad.
B. Entrecruzamiento.
C. Grupos.
D. No inferioridad.
E. Equivalencia.

9-12. En un ensayo aleatorizado controlado de dos fármacos para la enfermedad arterial coronaria, el resultado principal está compuesto por el infarto agudo de miocardio, la angina de pecho grave y la muerte por causa cardíaca. ¿Cuál de las siguientes es la ventaja principal de este abordaje?

A. Hay más episodios de resultados de los que podría haber para cualquiera de los resultados individuales.
B. Todos los resultados son igualmente afectados por las intervenciones.
C. El ensayo tiene más generalización.
D. Cada uno de los resultados individuales es importante por sí mismo.
E. Si uno de los resultados es infrecuente, los otros pueden compensarlo.

9-13. En un ensayo aleatorizado controlado que compara dos abordajes para tratar a los niños con bronquiolitis, las características iniciales de 200 niños en el ensayo son algo diferentes en los dos grupos asignados al azar. ¿Cuál de los siguientes podría explicar este hallazgo?

A. «Mala suerte» durante la aleatorización.

booksmedicos.org

B. Una revelación en el ocultamiento de la asignación.
C. Ambas.
D. Ninguna.

9-14. ¿Cuál de lo siguiente generalmente se aprende durante un ensayo de fase III sobre fármacos?

A. La relación entre la dosis y la eficacia.
B. Las tasas de efectos adversos infrecuentes.
C. La eficacia o efectividad.
D. El rango de dosis que es bien tolerada.

9-15. ¿Cuál de las siguientes es la ventaja principal de los ensayos aleatorizados controlados sobre los estudios observacionales de los efectos del tratamiento?

A. Un número menor de problemas éticos.
B. Prevención de la confusión.
C. Se parecen al tratamiento habitual.
D. Respuesta más rápida.
E. Menos costosos.

Encontrará las respuestas en el apéndice A.

BIBLIOGRAFÍA

1. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, et al. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med* 2008; 358:2545–2559.
2. Allen C, Glasziou P, Del Mar C. Bed rest: a potentially harmful treatment needing more careful evaluation. *Lancet* 1999;354:1229–1233.
3. Hawkins G, McMahon AD, Twaddle S, et al. Stepping down inhaled corticosteroids in asthma: randomized controlled trial. *BMJ* 2003;326:1115–1121.
4. Lamb SE, Marsh JL, Hutton JL, et al. Mechanical supports for acute, severe ankle sprain: a pragmatic, multicentre, randomized controlled trial. *Lancet* 2009;373:575–581.
5. Dykes PC, Carroll DL, Hurley A, et al. Fall prevention in acute care hospitals. A randomized trial. *JAMA* 2010;304:1912–1918.
6. Carson JL, Terrin ML, Noveck H, et al. Liberal or restrictive transfusion in high-risk patients after hip surgery. *N Engl J Med* 2011;365:2453–2462.
7. Sprung CL, Annane D, Keh D, et al. Hydrocortisone therapy for patients with septic shock. *N Engl J Med* 2008;358:111–124.
8. Avins AL, Pressman A, Ackerson L, et al. Placebo adherence and its association with morbidity and mortality in the studies of left ventricular dysfunction. *J Gen Intern Med* 2010;25: 1275–

booksmedicos.org

1281.
9. The AIM-HIGH Investigators. Niacin in patients with low HDL cholesterol levels receiving intensive statin therapy. *N Engl J Med* 2011;365:2255–2267.
10. Feldman T, Foster E, Glower DG, et al. Percutaneous repair or surgery for mitral regurgitation. *N Engl J Med* 2011;364: 1395–1406.
11. Mitja O, Hayes R, Ipai A, et al. Single-dose azithromycin versus benzathine benzylpenicillin for treatment of yaws in children in Papua New Guinea: an open-label, non-inferiority, randomized trial. *Lancet* 2012;379:342–347.
12. Chrischilles EA, Pendergast JF, Kahn KL, et al. Adverse events among the elderly receiving chemotherapy for advanced non-small cell lung cancer. *J Clin Oncol* 2010;28:620–627.
13. Kramarz P, DeStefano F, Gargiullo PM, et al. Does influenza vaccination exacerbate asthma. Analysis of a large cohort of children with asthma. Vaccine Safety Datalink Team. *Arch Fam Med* 2000;9:617–623.
14. Cates CJ, Jefferson T, Rowe BH. Vaccines for preventing influenza in people with asthma. Available at http://summaries.cochrane.org/CD000364/vaccines-for-preventing-influenza-in-people-with-asthma. Accessed July 26, 2012.
15. Feinstein AR, Horwitz RI. Double standards, scientific methods, and epidemiologic research. *N Engl J Med* 1982;307: 1611–1617.

booksmedicos.org

Capítulo 10

Prevención

Si un paciente le pide ayuda a un médico, este hace lo mejor que puede. Él no es responsable de los defectos en el conocimiento médico. Sin embargo, si el médico realiza procedimientos de tamizaje, está en una situación muy diferente. Debería tener evidencia concluyente que ese tamizaje puede modificar la historia natural de la enfermedad en una proporción significativa de esas personas examinadas.

—Archie Cochrane y Walter Holland
1971

PALABRAS CLAVE

Atención preventiva
Vacunaciones
Tamizaje
Orientación conductual
Quimioprevención
Prevención primaria
Prevención secundaria
Prevención terciaria
Vigilancia
Tamizaje de prevalencia
Tamizaje de incidencia
Sesgo de anticipación diagnóstica
Sesgo de duración
Sesgo de cumplimiento
Adherencia al placebo
Carcinomas de transición (de intervalo)
Método de detección

booksmedicos.org

Método de incidencia
Resultado falso positivo de la prueba de tamizaje
Efecto de etiquetado
Sobrediagnóstico
Enfermedad preexistente
Incidentaloma
Años de vida ajustados por calidad (AVAC)
Análisis costoefectividad

A la mayoría de los médicos les atrae su profesión porque esperan poder curar enfermedades. Con todo, la mayoría de la gente preferiría no contraer nunca enfermedad alguna o, si no pueden evitarla, prefieren que se les detecte pronto para erradicarla completamente antes de que les perjudique. Para lograr este objetivo, hay personas sin dolencias específicas que se someten a intervenciones con el fin de identificar y modificar los factores de riesgo, evitar el inicio de la enfermedad o detectarla al principio de su curso para que el tratamiento temprano la evite. Cuando estas intervenciones se realizan en la práctica clínica, la actividad se denomina **atención preventiva.**

La atención preventiva constituye una gran parte de la práctica clínica (1). Los médicos deben comprender la base conceptual y el contenido. Deben estar preparados para responder preguntas de los pacientes como: «Doctor, ¿cuánto ejercicio necesito?», o «He escuchado que un estudio demostró que los antioxidantes no son útiles para prevenir la enfermedad cardíaca, ¿qué piensa usted?» o «He visto un anuncio en el periódico sobre una prueba de calcio, ¿cree que debería hacerme una?».

Gran parte del enfoque científico para la prevención en la medicina clínica, especialmente los principios básicos de riesgo, el uso de pruebas diagnósticas, el pronóstico de enfermedad y la efectividad de las intervenciones, ya se ha abordado en este libro, pero en este capítulo ampliaremos aquellos principios y estrategias que se relacionan específicamente con la prevención.

ACTIVIDADES PREVENTIVAS EN LOS ÁMBITOS CLÍNICOS

En el ámbito clínico, las actividades de atención preventiva a menudo se pueden incorporar a la atención en curso de los pacientes como cuando el médico comprueba la presión arterial de un paciente que refiere dolor de

booksmedicos.org

garganta o recomienda la vacunación contra el neumococo en una persona mayor después de tratarlo por una erupción cutánea. Otras veces se conciertan visitas sólo para atención preventiva; así surgen los términos examen de salud periódico o reconocimiento médico anual.

Tipos de prevención clínica

Existen cuatro grandes tipos de atención clínica preventiva: vacunaciones, tamizaje (detección sistemática), orientación conductual (algunas veces hace referencia a cambios en el estilo de vida) y la quimioprofilaxis. Todos ellos se aplican a lo largo de la vida.

Vacunación

La **vacunación** infantil previene 15 enfermedades diferentes, determina en gran medida la programación de las consultas con el pediatra en los primeros meses de vida. Recientemente, se han añadido las vacunas contra el virus del papiloma humano (VPH) para niñas adolescentes para la prevención del cáncer de cuello uterino. Las vacunas para adultos incluyen los refuerzos para la difteria, la tos ferina y el tétanos (DPT), además de las vacunas para prevenir la influenza, la neumonía por neumococo y la hepatitis A y B.

Tamizaje

El **tamizaje** es la identificación de la enfermedad asintomática o de los factores de riesgo. Las pruebas de detección comienzan en el período prenatal (como el tamizaje para el síndrome de Down en los fetos de mujeres embarazadas mayores) y continúan a lo largo de toda la vida (p. ej., cuando se pregunta sobre la audición de los ancianos). La segunda mitad de este capítulo trata de los principios científicos del tamizaje.

Orientación conductual (cambios del estilo de vida)

Los médicos pueden proporcionar una **orientación conductual** eficaz para motivar cambios en el estilo de vida. Los médicos aconsejan a los pacientes dejar de fumar, hacer una dieta saludable, beber alcohol con moderación, hacer ejercicio y mantener relaciones sexuales seguras. Es importante tener pruebas de que (i) el cambio en la conducta disminuye el riesgo para el trastorno de interés y (ii) la orientación conduce a un cambio en la conducta antes de invertir tiempo y esfuerzo en este método de

booksmedicos.org

prevención (v. apartado siguiente).

Quimioprevención

La **quimioprevención** consiste en el uso de fármacos para evitar las enfermedades. Se utiliza en un período temprano de la vida (p. ej. la ingesta de folato durante el embarazo para prevenir los defectos de tubo neural y la profilaxis ocular antibiótica en todos los recién nacidos para prevenir la conjuntivitis gonocócica del recién nacido), pero también es frecuente en los adultos (p. ej., el ácido acetilsalicílico en la profilaxis del infarto de miocardio y el tratamiento con estatinas para la hipercolesterolemia).

NIVELES DE PREVENCIÓN

El diccionario Webster define prevención como «el acto de prevenir o impedir», y «cómo evitar que pase alguna cosa» (2). Con estas definiciones, casi todas las actividades de la medicina se podrían definir como prevención. Al fin y al cabo, el objetivo de los esfuerzos médicos consiste en prevenir la muerte, la enfermedad, la incapacidad, el malestar y la insatisfacción (comentado en el capítulo 1). Sin embargo, en la medicina clínica, la definición de prevención suele limitarse a las intervenciones en las personas que se sabe que no tienen la enfermedad particular de interés. Se han definido tres niveles de prevención: primaria, secundaria y terciaria (fig. 10-1).

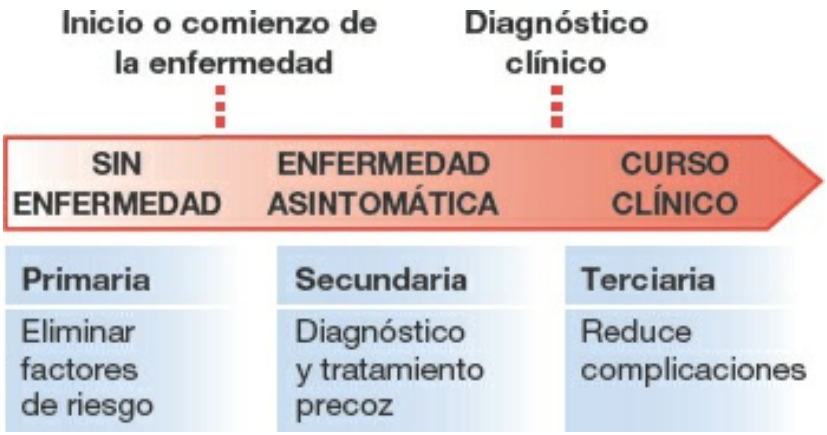


Figura 10-1 ■ Niveles de prevención. La prevención primaria evita la aparición de la enfermedad. La prevención secundaria detecta la enfermedad y la cura en la fase asintomática, y la prevención terciaria reduce las complicaciones de la enfermedad.

booksmedicos.org

Prevención primaria

La **prevención primaria** evita completamente la enfermedad, eliminando las causas. Las actividades clínicas preventivas más habituales en la atención primaria comprenden la vacunación para prevenir las enfermedades transmisibles, los fármacos y el asesoramiento conductual. Recientemente, la cirugía profiláctica se ha vuelto más habitual, con la cirugía bariátrica para prevenir las complicaciones de la obesidad, y la ooforectomía para prevenir el cáncer de ovario y la mastectomía, para prevenir el cáncer de ovario y de mama en mujeres con ciertas mutaciones genéticas.

La prevención primaria ha eliminado muchas enfermedades infecciosas de la infancia. En los hombres estadounidenses, esta prevención ha evitado muchas muertes por dos de las principales causas: el cáncer de pulmón y las enfermedades cardiovasculares. La mortalidad por cáncer de pulmón en los hombres disminuyó un 25 % de 1991 a 2007, estimando que se evitaron unas 250 000 muertes (3). Esta disminución se mantuvo de acuerdo con las tendencias para dejar de fumar entre los adultos, sin un tamizaje organizado y sin mucha mejoría en la supervivencia tras el tratamiento para el cáncer de pulmón. Las tasas de mortalidad por enfermedad cardíaca en los hombres han disminuido a la mitad en las últimas décadas (4), no sólo porque la atención médica ha mejorado, sino también debido a los esfuerzos de prevención primaria, como dejar de fumar y el uso de antihipertensivos y estatinas. Ahora es posible iniciar la prevención primaria para el cáncer de cuello uterino, hepatocelular, de piel y de mama, para las fracturas óseas y para el alcoholismo.

Un atributo especial de la prevención primaria que implica esfuerzos para ayudar a los pacientes a adoptar estilos de vida saludables es que una sola intervención puede prevenir varias enfermedades. Dejar de fumar no disminuye sólo el cáncer de pulmón, sino muchas otras enfermedades pulmonares, otros tipos de cáncer y, sobre todo, la enfermedad cardiovascular. Mantener un peso adecuado previene la diabetes y la osteoartritis, así como la enfermedad cardiovascular y algunos tipos de cáncer.

La prevención primaria a nivel comunitario también puede ser efectiva. Los ejemplos incluyen los requisitos de vacunación para estudiantes, las normas antitabaco en los edificios públicos, la cloración y fluorización

booksmedicos.org

(fluoración) de las fuentes de agua, y las leyes que exigen el uso de cinturón de seguridad en los automóviles y del casco en motocicletas y bicicletas. Algunas actividades de prevención primaria se llevan a cabo en ámbitos laborales específicos (uso de tapones para los oídos o mascarillas contra el polvo), en escuelas (vacunación) o en ámbitos sanitarios especializados (uso de pruebas para detectar la hepatitis B o el VIH en bancos de sangre).

Para algunos problemas, como las lesiones por accidentes automovilísticos, la prevención comunitaria funciona mejor. Para otros, como la profilaxis en los recién nacidos para prevenir la conjuntivitis gonocócica del recién nacido, el ámbito clínico funciona mejor. Aun para otras, los esfuerzos clínicos pueden complementar las actividades en toda la comunidad. En los esfuerzos para la prevención del tabaquismo, los médicos ayudan a pacientes individuales a dejar de fumar mientras que la educación pública, las nomas y los impuestos evitan que los adolescentes empiecen a fumar.

Prevención secundaria

La **prevención secundaria** detecta la enfermedad incipiente cuando es asintomática y cuando el tratamiento puede evitar que avance. Tiene un proceso de dos pasos que implica una prueba de tamizaje, un diagnóstico de seguimiento y un tratamiento para aquellos con la enfermedad de interés. Algunos ejemplos son el tamizaje para pacientes asintomáticos con VIH y los frotis rutinarios de Papanicolaou. La mayoría de la prevención secundaria se realiza en el ámbito clínico.

Como ya se ha señalado antes, el tamizaje es la detección de una enfermedad o un factor de riesgo (sin identificar) mediante el interrogatorio (p. ej., preguntar al paciente si es fumador), la exploración física (p. ej., la medida de la presión arterial), las pruebas de laboratorio (p. ej., la determinación de proteinuria en un diabético) u otros procedimientos (p. ej., la exploración de la densidad mineral ósea) que se pueden aplicar con razonable rapidez a las personas asintomáticas. Las pruebas de detección clasifican a las personas que supuestamente están bien (en cuanto a la enfermedad de interés) y que presentan una enfermedad o un factor de riesgo de una enfermedad, de aquellas que tienen una baja posibilidad. Estas pruebas forman parte de todas las acciones de prevención secundaria y algunas de las de atención primaria y terciaria. No

booksmedicos.org

pretenden ser diagnósticas. Si médico y paciente no han insistido en la investigación de los resultados anormales y el tratamiento, en caso de que sea necesario, las pruebas de detección no se deben realizar.

Prevención terciaria

La**prevención terciaria** hace referencia a las actividades clínicas que impiden un deterioro o reducen las complicaciones después de que la enfermedad se haya manifestado. Un ejemplo es el uso de β-bloqueadores para disminuir el riesgo de muerte en los pacientes que se han recuperado de un infarto de miocardio. La prevención terciaria es sólo otro término para tratamiento, pero enfocado en los efectos sobre la salud que se producen no tanto en horas y días, sino en meses y años. Por ejemplo, en pacientes diabéticos, un buen tratamiento requiere no únicamente el control de la glucosa sérica. La búsqueda y el tratamiento exitoso de otros factores de riesgo cardiovascular (p. ej., hipertensión, hipercolesterolemia, obesidad y tabaquismo) ayudan a prevenir la enfermedad cardiovascular en pacientes diabéticos tanto, o incluso más, que un buen control de la glucosa sérica. Además, los diabéticos requieren exploraciones oftalmológicas regulares para detectar la retinopatía diabética temprana, la atención rutinaria de los pies y la revisión de las proteínas urinarias para dirigir el uso de inhibidores de la enzima convertidora de angiotensina para prevenir la insuficiencia renal. Todas estas actividades preventivas son terciarias en el sentido de que previenen y reducen las complicaciones de una enfermedad que ya está presente.

Confusión sobre la prevención primaria, secundaria y terciaria

A lo largo de los años, a medida que una gran parte de la práctica clínica ha implicado la prevención, las diferencias entre la prevención primaria, secundaria y terciaria se han enmascarado. Históricamente, se pensaba que la prevención primaria era sobre todo la vacunación para las enfermedades infecciosas y el consejo sobre los hábitos de vida saludables, pero la prevención primaria ahora incluye prescribir antihipertensivos y estatinas para prevenir las enfermedades cardiovasculares y realizar cirugías profilácticas para prevenir el cáncer de ovario en mujeres con ciertas anomalías genéticas. Cada vez más, los factores de riesgo se tratan como si fueran enfermedades, incluso en un momento en que no han causado

booksmedicos.org

ninguna de estas. Esto es cierto para un número creciente de riesgos para la salud, por ejemplo, la densidad mineral ósea baja, la hipertensión, la hiperlipidemia, la obesidad y ciertas anomalías genéticas. El tratamiento de los factores de riesgo como enfermedades amplía la definición de prevención secundaria al ámbito de la prevención primaria tradicional.

En algunas disciplinas, como en cardiología, el término prevención secundaria se usa cuando se discute sobre la prevención terciaria. Se declaró «una nueva era de prevención secundaria» cuando se trató a los pacientes con síndrome coronario agudo (infarto de miocardio o angina inestable) con una combinación de tratamientos con antiagregantes plaquetarios y anticoagulantes para prevenir la muerte por causa cardiovascular (5). De forma similar, «la prevención secundaria» de la enfermedad cerebrovascular se utiliza para describir las intervenciones con el fin de prevenirla en pacientes con ataques isquémicos transitorios.

Las pruebas que se utilizan para la prevención primaria, secundaria y terciaria, así como para el diagnóstico, a menudo son idénticas, otra de las razones para confundir los niveles de prevención (y confundir prevención con diagnóstico). La colonoscopia se puede utilizar para encontrar un cáncer en un paciente con sangre en heces (diagnóstico); para hallar un cáncer de colon asintomático temprano (prevención secundaria); para extirpar un pólipo adenomatoso, que es un factor de riesgo para cáncer de colon (prevención primaria), o para revisar una recurrencia del cáncer en un paciente tratado por cáncer de colon (una actividad de prevención terciaria referida a menudo como **vigilancia**).

Independientemente de los términos que se utilicen, una razón subyacente para diferenciar los niveles entre las actividades preventivas es que existe un espectro de probabilidades de la enfermedad y los efectos adversos para la salud de la(s) enfermedad(es) que se están buscando y tratando durante las actividades preventivas, así como las diferentes probabilidades de efectos adversos para la salud derivados de las intervenciones utilizadas para la prevención en los diferentes niveles. El riesgo subyacente de ciertos problemas de salud suele ser mucho mayor en personas enfermas que en sanas. Por ejemplo, el riesgo de enfermedad cardiovascular en diabéticos es mucho mayor que en personas asintomáticas sin diabetes. Las pruebas idénticas actúan de forma diferente dependiendo del nivel de prevención. Además, las disyuntivas entre la efectividad y los daños pueden ser muy diferentes para los pacientes en

booksmedicos.org

diferentes partes del espectro. Los resultados falsos positivos de la prueba y el sobrediagnóstico (ambos se comentan más adelante en este capítulo) entre las personas sin la enfermedad que se está explorando son cuestiones importantes en la prevención secundaria, pero son menos importantes en el tratamiento de los pacientes que ya se sabe que tienen la enfermedad en cuestión. Los términos prevención primaria, secundaria y terciaria son formas para considerar estas diferencias conceptualmente.

ENFOQUE CIENTÍFICO DE LA PREVENCIÓN CLÍNICA

Cuando se considera qué actividades preventivas se deben realizar, el médico debe decidir, en primer lugar, cuáles son los problemas médicos o enfermedades que se deben intentar evitar. Esto es algo tan claro y evidente que parece innecesario mencionarlo; sin embargo, el hecho es que muchos procedimientos preventivos, especialmente las pruebas de detección sistemática, se realizan sin un conocimiento firme de lo que se está buscando o previniendo. Por ejemplo, los médicos que efectúan reconocimientos periódicos a sus pacientes, suelen pedir análisis de orina. Sin embargo, estas pruebas pueden utilizarse para detectar muchos otros problemas médicos, como la diabetes, las infecciones urinarias asintomáticas, el cáncer renal y la insuficiencia renal. Es necesario decidir cuál, si es que hay alguna, de estas afecciones se debe detectar antes de realizar la prueba. Uno de los avances científicos más importantes en la prevención clínica ha sido el desarrollo de métodos para decidir si una actividad preventiva propuesta debe llevarse a cabo (6). El resto de este capítulo describe estos métodos y conceptos.

Tres criterios son importantes cuando se determina si una enfermedad se debe incluir en la atención preventiva (tabla 10-1):

1. La magnitud del sufrimiento que provoca la enfermedad.
2. La efectividad, seguridad y costo de la intervención o el tratamiento preventivo.
3. El rendimiento de la prueba de tamizaje.

booksmedicos.org

Tabla 10-1

Criterios para decidir si una enfermedad se debe incluir en el cuidado preventivo

1. Cuál es la magnitud de la carga de sufrimiento provocada por la enfermedad en términos de:

Muerte	Malestar
Enfermedad	Insatisfacción
Discapacidad	Indigencia

2. Calidad de la prueba de detección, si se ha de realizar alguna, en términos de:

Sensibilidad	Seguridad
Especificidad	Aceptabilidad
Simplicidad	
Costo	

3. A. En la prevención primaria, cómo es la calidad de la intervención en términos de:

   Efectividad
   Seguridad
   Relación costo-efectividad

   o bien,

   B. En la prevención secundaria, si se detecta la enfermedad, cómo es la calidad del tratamiento que se administra en términos de:

   Efectividad
   Seguridad
   El tratamiento precoz tras el tamizaje es más eficaz que el tratamiento posterior sin tamizaje cuando el paciente comienza a tener síntomas
   Relación costo-efectividad

MAGNITUD DEL SUFRIMIENTO

Sólo las enfermedades que amenazan la vida o la salud (v. los resultados de la enfermedad del cap. 1) se deben incluir en la atención preventiva. La gravedad de la enfermedad está determinada, fundamentalmente, por (i) la magnitud del sufrimiento que causa a los afectados por la enfermedad y (ii) su frecuencia.

¿Cómo se mide el sufrimiento? A menudo se mide por las tasas de mortalidad y la frecuencia de hospitalizaciones, además de la cantidad de uso de los recursos sanitarios causados por la enfermedad. Se dispone de mucha menos información sobre cuánta discapacidad, dolor, náusea o

booksmedicos.org

insatisfacción causa una enfermedad determinada.

La frecuencia del trastorno también es importante para decidir sobre la prevención. Una enfermedad puede causar mucho sufrimiento a las personas que desgraciadamente la padecen, pero puede producirse con tanta infrecuencia (especialmente en el grupo de edad concreto del enfermo) como para considerar el tamizaje. El cáncer de mama es un ejemplo. Aunque puede ocurrir en mujeres mucho más jóvenes, la mayoría se presenta en mujeres mayores de 50 años. En una mujer de 20 años, la incidencia del cáncer de mama es de aproximadamente 1,6 por cada 100 000 (una quinta parte del índice de los hombres de poco más de 70 años) (7). Aunque el cáncer de mama se debe contemplar en la asistencia preventiva de las mujeres mayores, es demasiado infrecuente en mujeres de 20 años y en hombres de 70 como para que se realice el tamizaje. La detección sistemática de enfermedades muy raras significa no sólo que, como mucho, se beneficiarán de ella muy pocas personas, sino que también da lugar a pruebas con falsos positivos en algunas personas que pueden continuar sufriendo complicaciones de una evaluación diagnóstica posterior.

La incidencia de lo que se va a prevenir es especialmente importante en la prevención primaria y secundaria porque, independientemente de la enfermedad, el riesgo es bajo para la mayoría de las personas. La estratificación de la población de acuerdo con el riesgo y la selección de los grupos de alto riesgo puede ayudar a superar este problema, una práctica que se realiza con frecuencia al dirigir actividades preventivas específicas a ciertos grupos de edad.

EFECTIVIDAD DEL TRATAMIENTO

Como se destacó en el capítulo 9, los ensayos aleatorizados controlados son la evidencia científica más sólida para establecer la efectividad de los tratamientos. Es una práctica habitual cumplir con esta norma para la prevención terciaria (tratamientos). Por otro lado, para realizar ensayos aleatorizados en la evaluación de la prevención primaria o secundaria se requieren estudios muy grandes de miles, a menudo decenas de miles de pacientes, realizados durante varios años, a veces décadas, porque el resultado de interés es raro y a menudo tarda años en presentarse. La tarea es enorme; todas las dificultades de los ensayos controlados contempladas

booksmedicos.org

en el capítulo 9 se magnifican varias veces.

Otros desafíos en la evaluación de los tratamientos en la prevención se describen en el siguiente texto para cada nivel de prevención.

Tratamiento en la prevención primaria

Cualquiera que sea la prevención primaria (vacunas, fármacos, asesoramiento conductual o cirugía profiláctica) debe ser eficaz (capaz de producir un resultado beneficioso en situaciones ideales) y efectivo (capaz de producir un resultado beneficioso neto en condiciones habituales, considerando el cumplimento del paciente). Debido a que las intervenciones para la prevención primaria suelen aplicarse en un gran número de personas sanas, también deben ser muy seguras.

Ensayos aleatorizados

Virtualmente, todas las vacunas recomendadas están respaldadas por la evidencia obtenida de ensayos aleatorizados, algunas veces con relativa rapidez cuando los resultados ocurren en semanas o meses, como en las infecciones de la infancia. Debido a que los productos farmacéuticos están regulados, las actividades de prevención primaria y secundaria relacionadas con fármacos (p. ej., el tratamiento de la hipertensión y la hiperlipidemia en los adultos) también se han evaluado en ensayos aleatorizados. Cuando la prevención propuesta no está regulada, es menos habitual contar con ensayos aleatorizados, como ocurre con las vitaminas, minerales y suplementos alimenticios, o cuando la intervención es el asesoramiento conductual.

Estudios observacionales

Los estudios observacionales pueden ayudar a aclarar la efectividad de la prevención primaria cuando no es posible la aleatorización.

Ejemplo

Los ensayos aleatorizados han encontrado que las vacunas contra el virus de la hepatitis B (VHB) son altamente efectivas para prevenir la hepatitis B. Debido a que el VHB es un factor de riesgo importante para el cáncer hepatocelular, uno de los cánceres más habituales y mortales a nivel mundial, se pensó que la vacunación contra el VHB sería útil para prevenir el

booksmedicos.org

cáncer de hígado. Sin embargo, el intervalo habitual entre la infección por VHB y el desarrollo de cáncer hepatocelular es de dos a tres décadas y no sería ético negar la intervención al grupo de control contra una enfermedad infecciosa, para determinar si décadas después también fue efectivo contra el cáncer. Se realizó un estudio en Taiwán, donde empezó la vacunación nacional en 1984. Se compararon las tasas de incidencia de cáncer hepatocelular entre las personas vacunadas al nacer entre 1984 y 2004 frente a aquellas nacidas entre 1979 y 1984 cuando no existía ningún programa de vacunación. (La comparación fue posible gracias a una minuciosa base de datos nacional sobre la salud en la isla). La incidencia de este cáncer disminuyó casi el 70 % entre las personas jóvenes los 20 años después de introducir la vacuna contra el VHB (8).

Como se destacó en el capítulo 5, los estudios observacionales son vulnerables a sesgos. La conclusión de que la vacuna contra el VHB previene el carcinoma hepatocelular es razonable desde una perspectiva biológica y por el resultado dramático. Estaría en un terreno aún más sólido si los estudios de otras poblaciones sometidas a la vacunación confirmaran los resultados del estudio de Taiwán. En el capítulo 12 se habla de la construcción de un caso de causalidad en ausencia de ensayos aleatorizados.

Seguridad

En el caso de las vacunas, la aparición de efectos adversos podría ser tan rara que sería poco probable que los ensayos aleatorizados los descubrieran. Una forma de estudiar este problema es rastrear enfermedades en grandes bases de datos de millones de pacientes y comparar la frecuencia de los efectos adversos relacionados con el momento de la vacunación entre los grupos en diferentes períodos de tiempo.

Ejemplo

El síndrome de Guillain-Barré (SGB) es un trastorno neurológico autoinmune raro y grave caracterizado por parálisis ascendente que puede afectar temporalmente a los músculos respiratorios, por lo que se requiere de intubación y apoyo ventilatorio. Se desarrolló una vacuna contra la influenza porcina en 1970 que se asoció con un fuerte aumento inesperado del número de casos de SGB y contribuyó a la suspensión del programa de vacunación. En 2009 se creó una vacuna contra un nuevo tipo de virus de la influenza A (H1N1) de origen porcino, y se requería de un método para rastrear la incidencia de SGB. Una forma de hacerlo fue utilizar el sistema de vigilancia de las bases de datos electrónicas de salud con más de 9 millones de personas. Comparando la incidencia de SGB hasta 6 semanas después de la vacunación contra

booksmedicos.org

la incidencia anterior (antecedente) de SGB en el mismo grupo de personas vacunadas, se estimó que el riesgo atribuible a la vacuna en el 2009 fue de cinco casos adicionales de SGB por un millón de vacunaciones. Aunque la incidencia de SGB aumentó después de la vacunación, el efecto en 2009 fue de alrededor de la mitad de lo observado en la década de 1970 (9). La estimación muy baja del riesgo atribuible en el 2009 es tranquilizadora. Si los episodios raros, que ocurren en sólo unas pocas personas por millón, se han de detectar casi en tiempo real, se requieren sistemas de vigilancia basados en la población. Aun así, las asociaciones encontradas en los sistemas de vigilancia son una evidencia débil para una relación casual debido a su naturaleza observacional y a que las bases de datos electrónicas a menudo no tienen información sobre todas las posibles e importantes variables de confusión.

Asesoramiento

Las leyes estadounidenses no solicitan una evidencia rigurosa sobre la efectividad de los métodos de asesoramiento conductual. Sin embargo, los médicos deben contar con la evidencia científica antes de incorporar el asesoramiento rutinario en la atención preventiva; el asesoramiento que no funciona es una pérdida de tiempo, cuesta dinero y puede dañar a los pacientes. La investigación ha demostrado que ciertos métodos de asesoramiento pueden ayudar a los pacientes a cambiar algunas conductas de salud. Los esfuerzos para dejar de fumar han abierto el camino para muchos ensayos aleatorizados que evalúan diferentes enfoques.

Ejemplo

El tabaquismo mata a unos 450 000 estadounidenses cada año. No obstante, ¿cuál es la prueba de que el hecho de aconsejar a los pacientes el abandono de este hábito hace que estos realmente lo dejen? ¿Hay algunas propuestas mejores que otras? Estas preguntas las formuló una comisión que analizó todos los estudios realizados sobre el abandono del tabaquismo, basándose en ensayos aleatorizados (10). Se encontraron 43 ensayos que analizaban el tiempo de asesoramiento y observaron una relación dosis-respuesta: mejorando el tiempo de contacto, mejoraba el índice de abstinencia (fig. 10-2). Asimismo, la comisión reveló que los ensayos aleatorizados demostraban que el hecho de incorporar farmacoterapia, como el bupropión (un fármaco que actúa a nivel central y que disminuye el deseo), vareniclina (un agonista del receptor de nicotina), chicles de nicotina, inhaladores nasales o parches fueron efectivos. La combinación del asesoramiento y la medicación (evaluados en 18 ensayos) aumentó aún más la tasa de abandono del tabaquismo. Por otro lado, no hubo efectos de la acupuntura, los ansiolíticos ni los β-bloqueadores.

booksmedicos.org

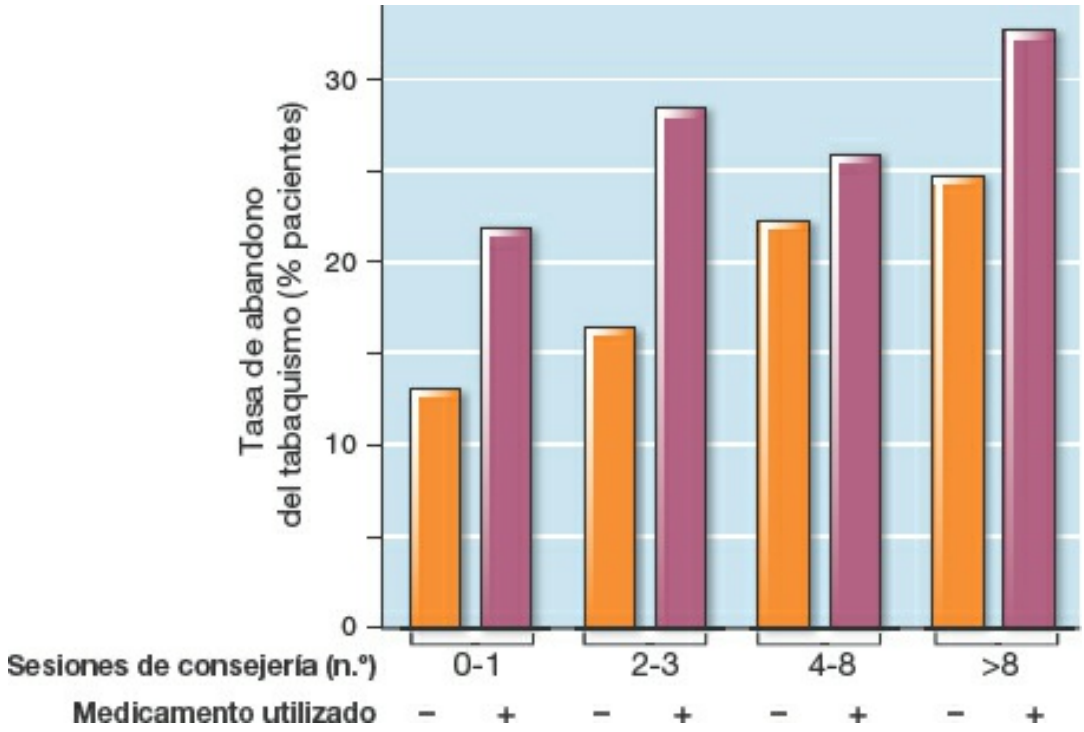


Figura 10-2 ■ Respuesta a la dosis de las tasas de abandono de acuerdo con el número de sesiones de asesoramiento que los médicos tienen con sus pacientes y el uso de medicamentos. (Datos de Fiore MC, Jaén CR, Baker TB, et al. Treating tobacco use and dependence: 2008 Update. Clinical Practice Guideline. Rockville, MD: U.S. Department of Health and Human Services. Public Health Service. May 2008.)

Tratamiento en la prevención secundaria

Los tratamientos para la prevención secundaria suelen ser los mismos que se utilizan en la medicina curativa. Al igual que las intervenciones para enfermedades sintomáticas, deberían ser eficaces y efectivos. Sin embargo, a diferencia de las intervenciones habituales para la enfermedad, en general se tarda años en establecer que una intervención de prevención secundaria es efectiva y se requiere de un gran número de personas para estudiarla. Por ejemplo, el tratamiento temprano posterior al tamizaje del cáncer colorrectal puede disminuir las muertes por esta causa en aproximadamente un tercio, pero para demostrar este efecto fue necesario un estudio de 45 000 personas con 13 años de seguimiento (11).

Un requisito único para el tratamiento en la prevención secundaria es que el tratamiento de la enfermedad *temprana* y *asintomática* debe ser superior al tratamiento de la enfermedad si se hubiera diagnosticado en el curso normal de los acontecimientos, cuando un paciente busca atención médica para sus síntomas. Si el resultado en las dos situaciones es el

booksmedicos.org

mismo, el tamizaje no añade ningún valor.

Ejemplo

El carcinoma pulmonar es la principal causa de muerte por cáncer en EE.UU. ¿Por qué, entonces, ningún gran grupo médico profesional recomienda el tamizaje para esta enfermedad a lo largo de la primera década del siglo XXI? Algunos ensayos aleatorizados que empezaron en los años 70 y en los 80 no consideraron el tamizaje preventivo. En un estudio sobre el uso de las radiografías de tórax y la citología de esputo para detectar el carcinoma pulmonar, se observó que los hombres fumadores a los que se cribaba cada 4 meses y se trataba rápidamente cuando se detectaba el carcinoma, no presentaron mejoría frente a los que no se les proponía el tamizaje y se les trataba sólo cuando aparecían los síntomas. Al cabo de 20 años, los índices de mortalidad eran los mismos en los dos grupos: 4,4 por 1 000 personas-años en los hombres a los que se les había realizado el tamizaje frente a 3,9 por 1 000 personas-años en hombres sin tamizar (12). Sin embargo, en 2011, un estudio aleatorizado controlado sobre la tomografía computarizada (TC) de dosis bajas para el tamizaje informó una reducción del 20 % en la mortalidad por carcinoma pulmonar después de una media de 6,5 años (13). Este ensayo demostró por primera vez de forma convincente que el tratamiento del carcinoma pulmonar asintomático detectado en el tamizaje disminuyó la mortalidad.

Tratamiento en la prevención terciaria

Todos los nuevos tratamientos farmacéuticos en EE.UU. están regulados por la Food and Drug Administration (FDA), que casi siempre solicita una prueba sobre la eficacia de ensayos clínicos aleatorizados. Es fácil asumir, por lo tanto, que los tratamientos de prevención terciaria se han evaluado cuidadosamente. Sin embargo, después de que un fármaco ha sido aprobado, puede ser utilizado para indicaciones nuevas que no se han evaluado. Los pacientes con algunas enfermedades tienen un mayor riesgo de otras enfermedades; por lo tanto, algunos fármacos se usan no sólo para tratar el trastorno para el que han sido aprobados, sino también para prevenir otras enfermedades para las que los pacientes tienen mayor riesgo. Distinguir entre los efectos terapéuticos probados de un medicamento para una cierta enfermedad y su efecto para prevenir otras enfermedades es un desafío sutil que afrontan los médicos cuando consideran las intervenciones de prevención terciaria que no se han evaluado para este propósito. Algunas veces, las evaluaciones cuidadosas han producido resultados sorprendentes.

booksmedicos.org

Ejemplo

En el capítulo 9 se presentó un ejemplo sobre un estudio de tratamiento para pacientes con diabetes tipo 2. Este mostraba un resultado sorprendente; el control estricto de la glucosa sérica (con un nivel por debajo del intervalo normal) no prevenía mejor la enfermedad cardiovascular y la mortalidad que el control más flexible. Los medicamentos para la diabetes utilizados en este y otros estudios que encontraron resultados similares se aprobaron después de estudios aleatorizados controlados. Sin embargo, los resultados que condujeron a la aprobación de los medicamentos no fueron sobre la prevención de la enfermedad cardiovascular a largo plazo (prevención terciaria) entre los diabéticos, sino del efecto de los fármacos sobre los niveles de glucosa sérica. Se pensó que estos medicamentos también disminuirían las enfermedades cardiovasculares porque los estudios observacionales habían mostrado que los niveles de glucosa sérica se correlacionaban con el riesgo de enfermedad cardiovascular. Al poner esta hipótesis a prueba en ensayos aleatorizados rigurosos se obtuvieron resultados sorprendentes e importantes. Como resultado, la prevención terciaria de la diabetes ha cambiado para incluir un tratamiento agresivo de otros factores de riesgo para enfermedad cardiovascular. Con este abordaje, un ensayo aleatorizado demostró que la enfermedad cardiovascular y la muerte en los diabéticos se redujeron en un 50 % durante 13 años, en comparación con el grupo que recibió el tratamiento convencional destinado a controlar los niveles de glucosa sérica (14).

PROBLEMAS METODOLÓGICOS AL EVALUAR LOS PROGRAMAS DE TAMIZAJE

Varios problemas surgen al evaluar los programas de tamizaje, algunos de los cuales pueden hacer que parezca que el tratamiento precoz es efectivo tras el tamizaje cuando no lo es. Entre ellos se encuentran la diferencia entre los tamizajes de prevalencia e incidencia, y tres sesgos que se pueden producir en los estudios de tamizaje: sesgos de anticipación diagnóstica, de duración y de cumplimiento.

Tamizajes de incidencia y de prevalencia

Con el tiempo, el rendimiento del tamizaje desciende, a medida que este se repite; en la figura 10-3 se demuestra por qué ocurre esto. La primera vez que se realiza el **tamizaje de prevalencia,** la enfermedad ya hace tiempo que está presente. Durante el segundo tamizaje, la mayoría de los casos detectados se habrán iniciado entre el primer y el segundo tamizaje. (Algunos habrán pasado desapercibidos en el primer tamizaje). Por lo tanto, el segundo (y posterior) tamizaje se denomina **tamizaje de incidencia.** Así, cuando se realizan prue bas de detección periódicas a un grupo de personas, el número de casos de la enfermedad presente en el

booksmedicos.org

grupo desciende tras el tamizaje de prevalencia. Esto significa que el valor predictivo positivo de los resultados de las pruebas disminuirá tras la primera serie de tamizaje.

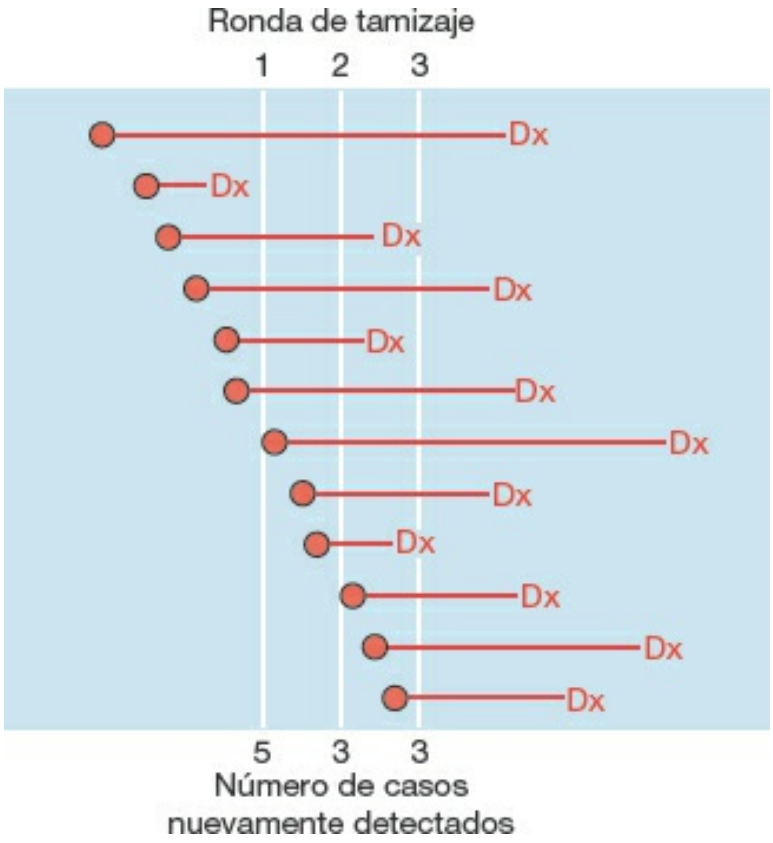


Figura 10-3 ■ Disminución del rendimiento de una prueba de detección tras la primera ronda de tamizaje. La primera ronda (tamizaje de prevalencia) detecta los casos prevalentes. La segunda y la tercera ronda (tamizajes de incidencia) detectan los casos incidentes. En esta figura se supone que la prueba detecta todos los casos y que se hace un tamizaje de toda la población. Cuando esto no es así, los casos que no se han detectado en la primera ronda se pueden detectar en las rondas posteriores (y el rendimiento será mayor). Dx, tiempo de diagnóstico, si no se realizara el tamizaje; O, inicio de la enfermedad.

booksmedicos.org

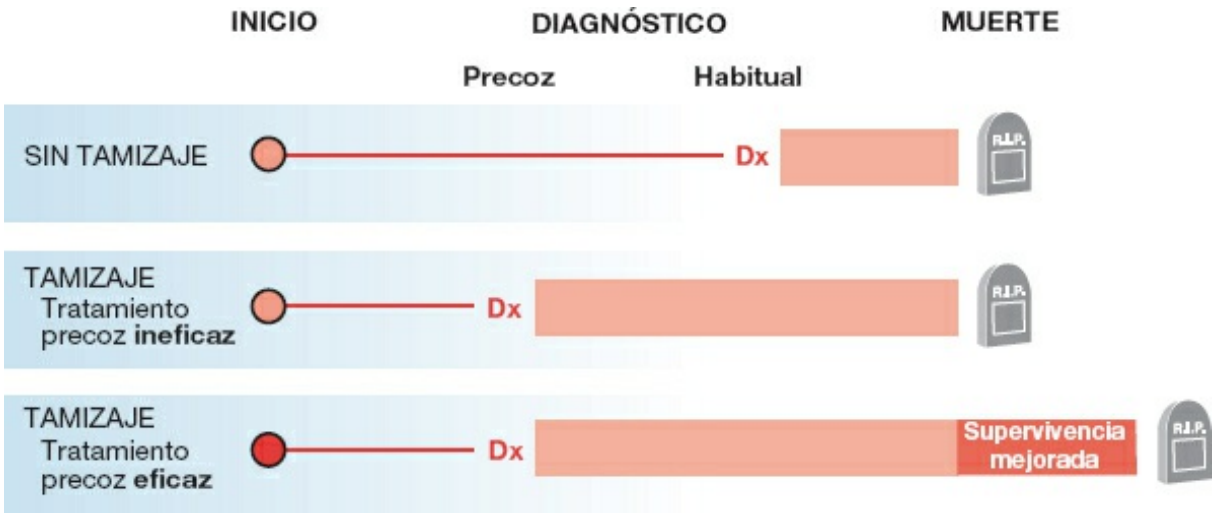


Figura 10-4 ■ Cómo afecta la anticipación diagnóstica al tiempo de supervivencia tras el tamizaje; O, inicio de la enfermedad. Las *zonas sombreadas* indican el tiempo de supervivencia tras el diagnóstico (Dx).

Sesgos especiales

Es más probable que los siguientes sesgos sean un problema en los estudios observacionales de tamizaje.

Sesgo de anticipación diagnóstica

La **anticipación diagnóstica** es el período de tiempo entre la detección por tamizaje de una enfermedad y el momento en que normalmente se diagnosticaría, porque un paciente presenta síntomas y solicita asistencia médica (fig. 10-4). La magnitud de anticipación diagnóstica de una determinada enfermedad está en función del índice biológico de progresión de la enfermedad y de la rapidez en que se detecte mediante el tamizaje.

Cuando la anticipación diagnóstica es muy breve, como sucede en el cáncer de pulmón, es difícil demostrar que el tratamiento de las condiciones médicas detectadas en el tamizaje es más efectivo que el tratamiento después que los síntomas hayan aparecido. Por otro lado, cuando hay una gran anticipación diagnóstica, como sucede en el cáncer de cuello uterino (aproximadamente, se tardan 20-30 años para que el carcinoma *in situ* desarrolle una enfermedad clínicamente invasiva), el tratamiento de la enfermedad que ha detectado el tamizaje puede ser muy eficaz.

¿Cómo la anticipación diagnóstica puede provocar resultados sesgados en un estudio sobre la eficacia del tratamiento precoz? Como se muestra en

booksmedicos.org

la figura 10-4, gracias al tamizaje, una enfermedad se puede detectar antes de lo que se habría detectado una vez que el paciente desarrollara los síntomas. Debido a ello, las personas a las que se diagnostica una enfermedad mortal mediante el tamizaje suelen sobrevivir más tiempo desde el momento del diagnóstico que los pacientes a los que se les diagnostica tras la aparición de síntomas, aunque el tratamiento no sea más eficaz que el del momento de la presentación clínica. En una situación así, puede parecer que el tamizaje ayuda a que los enfermos vivan más tiempo, aumentando los índices de supervivencia cuando, en realidad, no habrían obtenido más «tiempo de supervivencia», sino más «tiempo de enfermedad».

Tabla 10-2

Evitar sesgos en el tamizaje		
Sesgo	**Efecto**	**Cómo evitarlo**
Anticipación diagnóstica	Parece mejorar el tiempo de supervivencia, pero en realidad aumenta el tiempo de «enfermedad» tras la detección de esta	Utilizar las tasas de mortalidad en lugar de las de supervivencia
Duración	El resultado parece mejorar en el grupo tamizaje porque se detectan más cánceres con buen pronóstico	Comparar los resultados en RTC con un grupo de control y un grupo al que se le ha propuesto el tamizaje. Recuento de *todos* los resultados independientemente del método de detección
Cumplimiento	El resultado del grupo tamizaje parece ser mejor debido al cumplimiento, no a la prueba de	Comparar resultados en RTC con un grupo de control y uno al que se le ha propuesto el tamizaje. Recuento de *todos* los resultados

booksmedicos.org

	detección	independientemente del cumplimiento

ECA, ensayo clínico controlado aleatorizado.

Un método adecuado de análisis para evitar el sesgo de anticipación diagnóstica es comparar los índices de mortalidad de una edad específica en lugar de los índices de supervivencia en el grupo de personas que se sometieron al tamizaje y en un grupo control de personas similares que no se sometieron al tamizaje, como en un ensayo aleatorizado (tabla 10-2). Se sabe que el tamizaje para el cáncer de mama, de pulmón y colorrectal es eficaz porque los estudios han demostrado que los *índices de mortalidad* de las personas en las que se realiza un tamizaje son inferiores a las del grupo comparativo de personas no cribadas.

Sesgo de duración

El **sesgo de duración** se produce porque la proporción de las lesiones de crecimiento lento diagnosticadas mediante el tamizaje es mayor que la proporción de las diagnosticadas en la asistencia médica habitual. En consecuencia, el sesgo de duración hace que el tamizaje y el tratamiento precoz parezcan más eficaces que la asistencia habitual.

El sesgo de duración se produce del siguiente modo: el tamizaje es más eficaz cuando la enfermedad se desarrolla lentamente. Sin embargo, la mayoría de neoplasias malignas presentan un gran margen de índices de crecimiento. Algunas se desarrollan paulatinamente, y otras lo hacen muy rápido. Es probable que las pruebas de tamizaje detecten la mayoría de tumores de crecimiento lento, porque existen durante un período largo, antes de que sus síntomas se manifiesten. Es posible que los tumores de crecimiento rápido causen más síntomas, lo que conlleva al diagnóstico en el intervalo entre los exámenes de tamizaje (figs. 10-5 y 10-6). Por lo tanto, las pruebas de tamizaje tienden a detectar tumores con mejor pronóstico intrínseco, y debido a ello, los índices de mortalidad de las neoplasias malignas detectadas mediante el tamizaje pueden ser mejores que los de las que no se detectan mediante este procedimiento, aunque en esta situación, esa prueba no sea preventiva.

booksmedicos.org

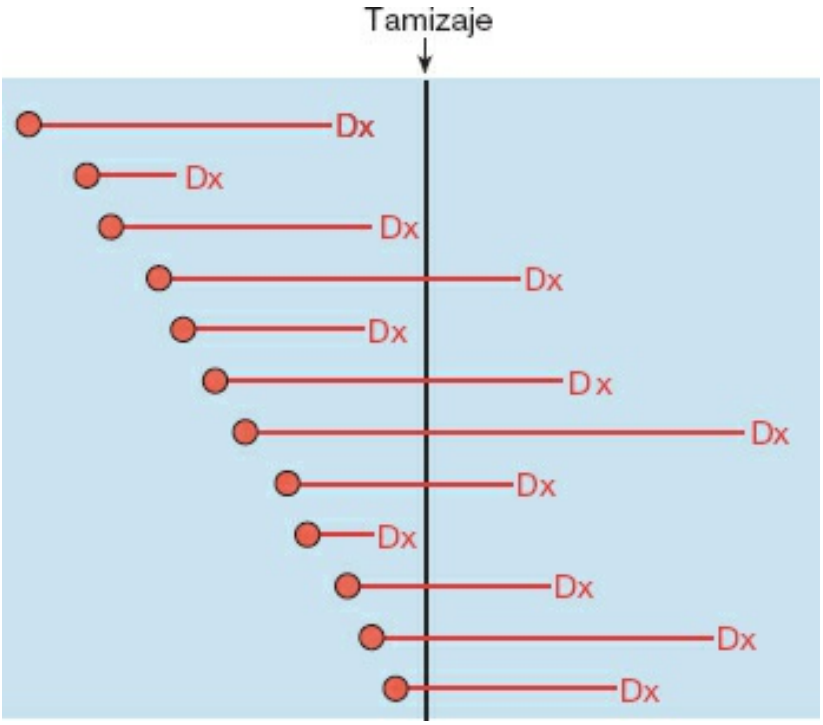


Figura 10-5 ■ **Sesgo de anticipación diagnóstica.** Es menos probable que los casos que evolucionan rápidamente desde el inicio (O) hacia los síntomas y el diagnóstico (Dx) se detecten durante una prueba de tamizaje.

Sesgo de cumplimiento

El tercer tipo más importante de sesgo que se puede producir en los estudios de prevención es el **sesgo de cumplimiento.** Los pacientes que respetan el cumplimiento tienden a tener mejor pronóstico, a pesar de las actividades de prevención. Las razones para ello no están completamente claras, pero en promedio, los pacientes que respetan el cumplimiento están más interesados en su salud y suelen estar más sanos que los que no lo respetan. Por ejemplo, un estudio aleatorizado que invitó a las personas para el tamizaje observó que los voluntarios del grupo control que no fueron invitados, pero solicitaron el tamizaje, tuvieron mejores tasas de mortalidad que el grupo de invitados, formado por personas que respetaron el cumplimiento y solicitaron el tamizaje, y por aquellos que respetaron el cumplimiento y rechazaron el tamizaje (15). El efecto del cumplimiento del paciente, a diferencia del efecto del tratamiento, ha comprendido principalmente la adherencia al medicamento en el grupo placebo, y se ha denominado **adherencia al placebo.**

booksmedicos.org

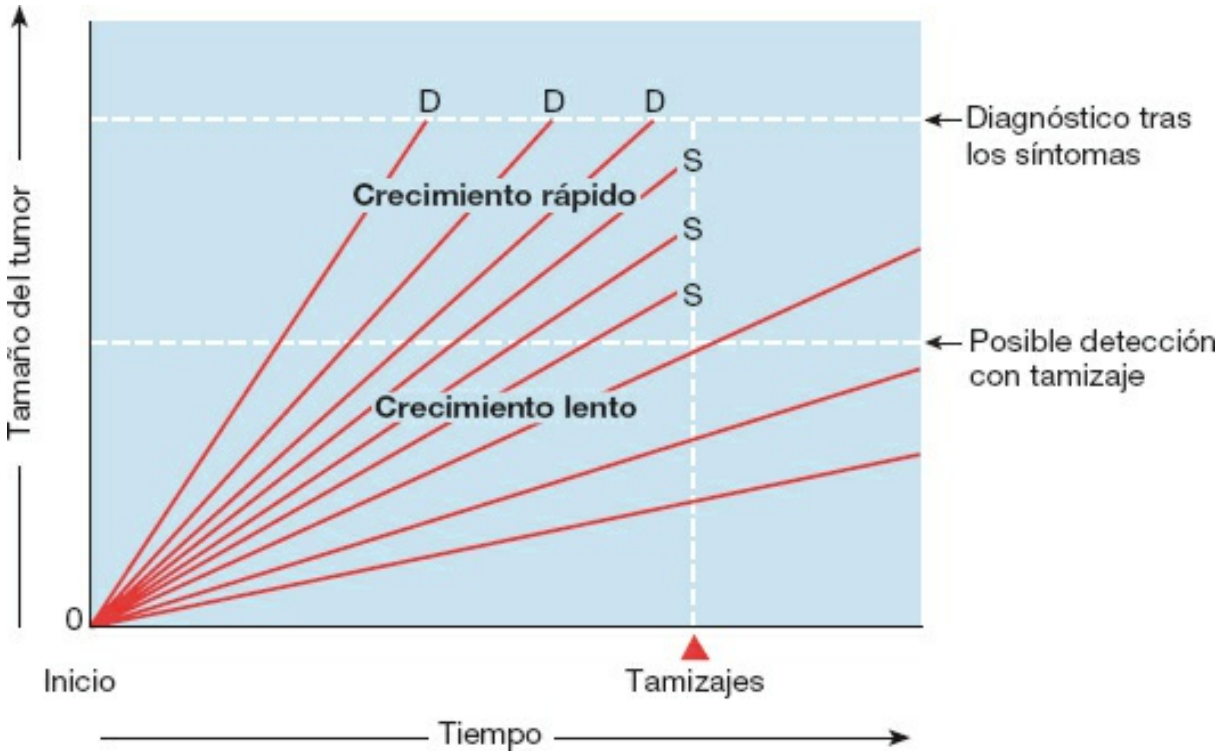


Figura 10-6 ■ Relación entre el sesgo de anticipación diagnóstica y la velocidad del crecimiento del tumor. Los tumores de crecimiento rápido llaman la atención médica antes de realizar el tamizaje; en cambio, los tumores de crecimiento más lento permiten más tiempo para la detección. D, diagnóstico tras los síntomas; S, detección tras el tamizaje.

Ejemplo

Se realizó un análisis para determinar si los resultados sobre la salud eran diferentes en el grupo placebo de un ensayo aleatorizado entre los pacientes que mostraron diferentes grados de adherencia al placebo (16). El estudio original fue un ensayo de prevención terciaria que asignó al azar a los pacientes asintomáticos con disfunción ventricular izquierda (fracción de eyección cardíaca < 35 %) para recibir el tratamiento activo (enalapril) o el placebo. El análisis demostró que después de 3 años, los pacientes aleatorizados al placebo que tomaban al menos el 75 % del fármaco placebo (pacientes con «alta adherencia») tenían la mitad de la tasa de mortalidad que los pacientes con «menor adherencia» aleatorizados al placebo. Se buscaron explicaciones para este extraño resultado. La diferencia en la mortalidad no cambió tras ajustar varios factores de riesgo, incluyendo las enfermedades serias. El medicamento placebo en sí era inerte químicamente. La adherencia al placebo parece ser un marcador de las personas más sanas aunque el mecanismo subyacente aún continúa sin determinarse.

Los sesgos debidos a la duración y al cumplimiento del paciente se pueden evitar con estudios que dependen de grupos de tamizaje y de control simultáneos que son comparables. Deben contarse todas las

booksmedicos.org

personas de cada grupo que presenten resultados de interés, independientemente del método de diagnóstico o grado de participación (tabla 10-2). Los estudios aleatorizados son el modelo más sólido, porque los pacientes asignados de forma aleatoria tendrán un número comparable de tumores de crecimiento rápido y lento y, por término medio, niveles comparables de cumplimiento. A continuación, para evitar los sesgos de anticipación diagnóstica, se puede realizar un seguimiento de estos grupos con respecto al índice de mortalidad, y no al de supervivencia. Si no es posible realizar un estudio aleatorizado, pueden ser válidos los resultados de los estudios basados en la población. En estos casos, la población tanto del grupo de tamizaje como del de control debe ser similar, la población de control no debe tener acceso al tamizaje, y se debe realizar un seguimiento minucioso de ambas poblaciones para documentar todos los casos de la evolución que se está estudiando.

Dado que tanto los ensayos controlados y aleatorizados como los estudios de población previstos son difíciles de llevar a cabo, duran mucho tiempo y son caros, a veces, los científicos intentan utilizar otros tipos de estudios para investigar los métodos preventivos, como los estudios de cohorte (v. cap. 5) o los estudios de casos y controles (v. cap. 6).

Ejemplo

Para comprobar si el tamizaje periódico con sigmoidoscopia reduce la mortalidad por cáncer colorrectal, los científicos investigaron la frecuencia de tamizaje con esta técnica durante los 10 años previos en pacientes que fallecieron por carcinoma y pacientes sanos, pareados por edad y sexo (17). Para abordar los sesgos de anticipación diagnóstica y de duración, investigaron el tamizaje sólo en personas que se sabía que habían fallecido por un cáncer colorrectal (grupo de casos) o que no habían fallecido por esta enfermedad (grupo de control). Para abordar los sesgos de cumplimiento, ajustaron los resultados al número de exámenes periódicos de salud generales que se llevaron a cabo con cada persona. También ajustaron los resultados para la presencia de enfermedades que podían haber provocado un aumento del tamizaje y un aumento de la posibilidad de cáncer colorrectal. Era menos probable que los pacientes que fallecieron a causa de esta afección en el recto o el sigmoides distal (donde los tumores se podrían haber detectado mediante la sigmoidoscopia) se hubieran hecho un tamizaje en los 10 años anteriores (8,8 %) en comparación con los pacientes del grupo de control (24,2 %). Los científicos observaron que la sigmoidoscopia junto con el tratamiento precoz evitaban casi el 60 % de las muertes a causa de neoplasias malignas colorrectales en el colon distal o en el recto, equivalente a una reducción del 30 % en la mortalidad por cáncer de colon y recto como un todo. Asimismo, al demostrar que no había protección alguna para este tipo de carcinomas por encima del nivel alcanzado con la sigmoidoscopia, los autores

booksmedicos.org

sugirieron que «es difícil concebir cómo esta especificidad anatómica del efecto se puede explicar mediante factores de confusión». Los hallazgos de este estudio han sido corroborados por dos ensayos aleatorizados subsecuentes sobre la sigmoidoscopia.

CRITERIOS PARA REALIZAR LAS PRUEBAS DE TAMIZAJE

Las pruebas utilizadas para el tamizaje deben cumplir con los criterios de las pruebas diagnósticas descritas en el capítulo 8.

Los siguientes criterios para realizar una buena prueba de tamizaje se aplican a todos los tipos de estas pruebas, ya sean datos obtenidos por interrogatorio, por exploraciones físicas o pruebas de laboratorio.

Sensibilidad y especificidad elevadas

La verdadera naturaleza de la búsqueda de una enfermedad en personas asintomáticas significa que la prevalencia normalmente es muy baja, incluso en grupos de riesgo elevado que se seleccionan por la edad, el sexo y otras características. Por lo tanto, una buena prueba de tamizaje debe tener una gran sensibilidad, para que no omita los pocos casos de enfermedad presentes. Además, debe ser sensible al inicio de la enfermedad, cuando todavía se puede modificar el curso posterior. La prueba de tamizaje es inútil si sólo es sensible en la fase avanzada de la enfermedad, cuando ya está demasiado adelantada para que el tratamiento sea eficaz. Asimismo, debe tener una gran especificidad para reducir la cantidad de personas con resultados falsos positivos que necesitan evaluación diagnóstica.

La sensibilidad y la especificidad son más decisivas en las pruebas de tamizaje que en las de diagnóstico, con una gran diferencia. Como se comentó en el capítulo 8, la sensibilidad y la especificidad de una prueba diagnóstica vienen determinadas por la comparación de sus resultados con los de otra prueba (el criterio de referencia o estándar de oro). En el tamizaje, el criterio de referencia de la presencia de una enfermedad no es sólo otra prueba, más precisa, sino también un *período de seguimiento.* El criterio de referencia sólo se realiza de forma sistemática a las personas cuyos resultados de la prueba de tamizaje son positivos, para diferenciar entre resultados falsos y verdaderos positivos. Se aplica un período de seguimiento a todos los individuos que presentan una prueba de tamizaje

booksmedicos.org

negativa, con objeto de diferenciar entre resultados verdaderos negativos y falsos negativos.

El seguimiento es especialmente importante en la detección sistemática del cáncer, ya que es entonces cuando surgen los **carcinomas de transición (de intervalo),** cánceres no detectados durante el tamizaje pero descubiertos posteriormente durante el período de seguimiento. Cuando se producen los carcinomas de transición, el cálculo de la prueba de sensibilidad es baja.

Ejemplo

En el capítulo 8 presentamos un estudio en el que se midieron los niveles de antígeno prostático específico (PSA) en muestras de sangre almacenadas obtenidas de una cohorte de hombres sanos (médicos) que más tarde fueron o no diagnosticados con cáncer de próstata (18). Trece de los 18 hombres diagnosticados con cáncer de próstata tenían 1 año después de la muestra de sangre niveles elevados de PSA (> 4,0 ng/ml) y habían sido diagnosticados después de un resultado anormal de PSA; los otros cinco tenían resultados anormales de PSA y desarrollaron carcinomas de transición durante el primer año tras una prueba normal de PSA. Por lo tanto, la sensibilidad del PSA se calculó como 13 dividido entre (13 + 5), o 72 %.

El objetivo clave consiste en escoger el período de seguimiento correcto. Si este es demasiado breve, puede que la enfermedad, que la prueba de tamizaje no ha detectado, no se manifieste de forma evidente, y se puede sobreestimar la sensibilidad de la prueba; por otro lado, si el período de seguimiento es demasiado largo, podría detectarse la enfermedad que no estaba presente en el momento del tamizaje, lo que comporta una baja estimación falsa de la prueba de sensibilidad.

Métodos de detección e incidencia para calcular la sensibilidad

El **método de detección** es un procedimiento que se utiliza para calcular la sensibilidad, tomando como verdaderos positivos a los casos que se detectan durante el tamizaje y a los carcinomas en transición como falsos negativos (tabla 10-3). Este método funciona bien en muchas **pruebas de detección,** pero en algunas pruebas de tamizaje de cánceres, existen dos dificultades. En primer lugar, como ya se ha mencionado, es necesario que se conozca la cantidad adecuada de tiempo de seguimiento para los carcinomas de transición; a menudo se ignora, y debe ser estimado. El

booksmedicos.org

método de detección supone también que las anomalías que ha detectado la prueba de tamizaje continuarán causando dificultades si se dejan. Este segundo punto constituye un problema en la detección de algunas neoplasias malignas, especialmente las de próstata.

Ejemplo

El cáncer de próstata histológico es frecuente en hombres, especialmente de edad avanzada. Una revisión de los estudios de autopsia mostró que en EE.UU., las tasas de prevalencia del cáncer de próstata oscilaron de un 8 % entre los hombres a sus 20 años hasta un 83 % entre los hombres a sus 70 años (19). Las pruebas de tamizaje pueden detectar este tipo de neoplasias malignas en muchos hombres, pero en la mayoría el carcinoma nunca será invasivo. Así, cuando la sensibilidad de las pruebas para el cáncer de próstata, como el PSA se determina por el método de detección, la prueba puede parecer bastante buena, porque el numerador incluye todos los cánceres detectados, no sólo aquellos con un potencial maligno.

Tabla 10-3

Cálculo de la sensibilidad de una prueba de detección de cáncer según el método de detección y el método de incidencia

Ejemplo teórico

Se introduce una nueva prueba de detección para el cáncer pancreático. En un grupo que se somete a tamizaje, se detecta cáncer en 200 personas; durante el siguiente año a otras 50 que habían obtenido resultados negativos del tamizaje, se les diagnostica carcinoma pancreático. En un grupo de control concurrente con las mismas características y el mismo tamaño, los miembros no se sometieron a un tamizaje; durante el año, se diagnosticó cáncer de páncreas a 100 personas

Sensibilidad de la prueba con el método de detección

$$\text{Sensibilidad} = \frac{\text{Número de cánceres detectados con tamizaje}}{\text{Número de cánceres detectados con tamizaje, más el número de carcinomas de transición}}$$

$$= \frac{200}{(200 + 50)}$$

$$= 0{,}80 \text{ o } 80\,\%$$

Sensibilidad de la prueba con el método de incidencia

$$\text{Sensibilidad} = 1 - (\text{tasa de intervalo en el grupo de tamizaje/tasa de incidencia del grupo de control})$$

$$= 1 - \left(\frac{50}{100}\right) = 0{,}50 \text{ o } 50\,\%$$

El **método de incidencia** calcula la sensibilidad utilizando la incidencia en personas a las que no se ha realizado el tamizaje y la tasa de carcinoma de transición en las personas a las que sí (tabla 10-3). El motivo de esta solución es que la sensibilidad de una prueba debe afectar a los índices de carcinoma de transición, pero no a la incidencia de la enfermedad. En el cáncer de próstata, el método de incidencia define la sensibilidad de la prueba como 1 menos la proporción de la tasa de

booksmedicos.org

carcinomas de próstata de transición en un grupo de hombres a los que se tamiza con regularidad, respecto a la incidencia del carcinoma de próstata de un grupo de hombres a los que no (grupo de control). El método de incidencia para calcular la sensibilidad soluciona el problema del recuento de neoplasias de próstata «benignas», pero puede subestimar la sensibilidad porque excluye los carcinomas de anticipación diagnóstica. Por consiguiente, la verdadera sensibilidad de una prueba probablemente se encuentre entre el cálculo de los dos métodos.

Valor predictivo positivo bajo

Debido a la baja prevalencia de la mayoría de las enfermedades en personas asintomáticas, el valor predictivo de un resultado positivo de la mayor parte de pruebas de tamizaje es bajo, incluso en las pruebas de especificidad elevada. (El caso opuesto se cumple para el valor predictivo negativo; porque cuando la prevalencia es baja, es probable que el valor predictivo negativo sea alto.) Los médicos que realizan pruebas de tamizaje a sus pacientes deben aceptar el hecho de que tendrán que estudiar a muchos pacientes con resultados de la prueba positivos, pero que no presentan la enfermedad. Sin embargo, al concentrar los esfuerzos del tamizaje en personas con una prevalencia mayor para la enfermedad, se puede reducir al mínimo el problema.

Ejemplo

La incidencia del cáncer de mama aumenta con la edad, desde aproximadamente 1 por cada 100 000 al año, a los 20 años, hasta 1 por cada 200 al año, pasados los 70. Además, la sensibilidad y la especificidad son mejores en las mujeres mayores. Por consiguiente, un nódulo detectado en el tamizaje en la mama de una mujer joven es más probable que no sea maligno que uno que se encuentre en una mujer mayor. En un gran estudio realizado sobre el tamizaje del cáncer de mama, la detección de carcinomas tras una mamografía anómala variaba considerablemente según la edad de la mujer; en las mujeres con poco más de 40 años, hubo unas 57 sin carcinomas; sin embargo, cuando se realizó en cada mujer un examen más detallado, se detectó una neoplasia maligna (fig. 10-7) con un valor predictivo positivo del 1,7 % (20). No obstante, en las mujeres de 80 años, el número descendía aproximadamente a 10, con un valor predictivo positivo del 9,5 %.

booksmedicos.org

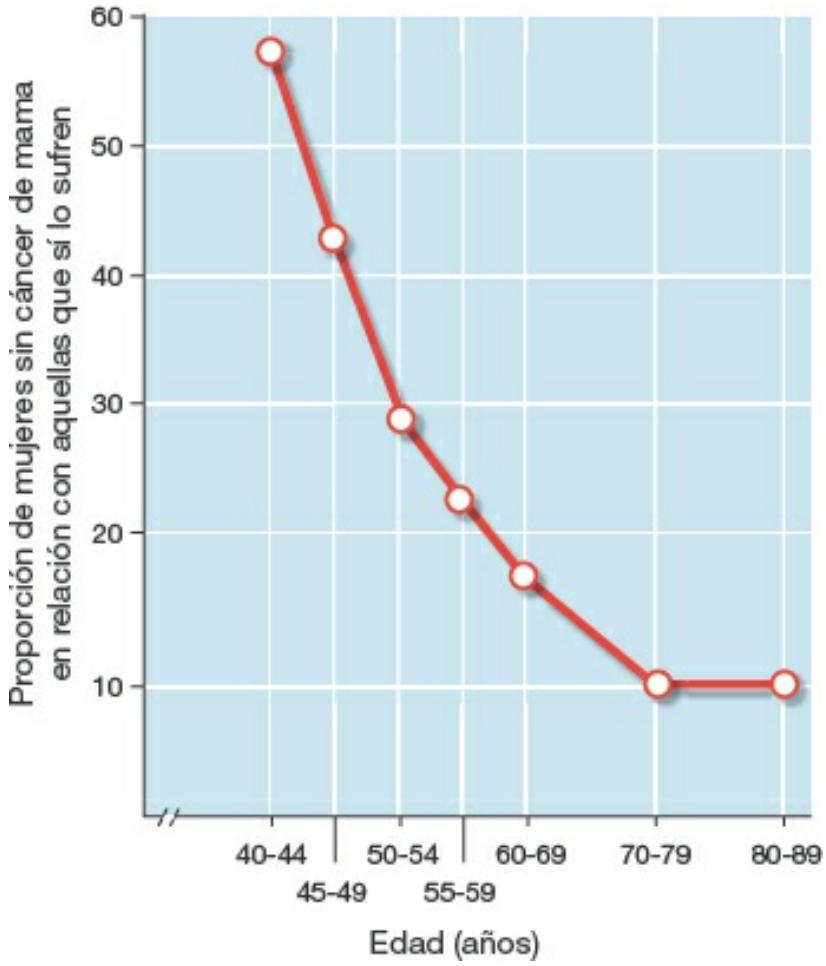


Figura 10-7 ■ Rendimiento de un tamizaje anormal de mamografías en función de la edad de las pacientes. Número de mujeres sin carcinoma mamario por cada mujer con un diagnóstico de esta enfermedad entre mujeres con mamografías anómalas cuando se les realizó el tamizaje para detectar cáncer de mama. (Datos de Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med 2003;138:168-175.)

Simplicidad y bajo costo

Una prueba de tamizaje ideal debe durar sólo unos pocos minutos, apenas preparación por parte del paciente, no depender de citas especiales y no ser cara.

Los exámenes sencillos y rápidos como las determinaciones de la presión arterial son tamizajes ideales; en cambio, pruebas como la colonoscopia, que es cara y debe planificarse y realizarse una preparación intestinal, son más apropiadas como pruebas de diagnóstico en pacientes con síntomas e indicaciones clínicas. Sin embargo, se ha encontrado que la colonoscopia de tamizaje puede ser muy eficaz para disminuir la

booksmedicos.org

mortalidad por cáncer colorrectal, y una prueba negativa no debe repetirse durante varios años. Otras pruebas, como la exploración del campo visual para detectar el glaucoma, y las audiometrías, para detectar la pérdida auditiva, se encuentran entre estos dos extremos.

El «costo» económico de la prueba no sólo depende del precio del procedimiento en sí y de los honorarios, sino también del costo de las evaluaciones posteriores que se realizan en pacientes con resultados positivos de las pruebas. Así, la sensibilidad, la especificidad y el valor predictivo afectan al costo. Además, también influye el hecho de si la prueba requiere una consulta especial del médico. Las pruebas de tamizaje realizadas cuando el paciente va al médico por otras causas (como sucede con las mediciones de la presión arterial) son mucho más económicas para los pacientes que las que requieren visitas especiales, tiempo extra fuera del horario de trabajo y transporte adicional. El precio también está determinado por la frecuencia con la que se repite la prueba. Conociendo todos estos puntos a la vez, se puede llegar a conclusiones sorprendentes.

Ejemplo

Se pueden utilizar varias pruebas diferentes para detectar carcinomas colorrectales: pruebas de detección de sangre oculta en heces, sigmoidoscopias cada 5 años y colonoscopias cada 10 años. El costo por adelantado de las pruebas varía considerablemente, oscila entre unos 20 dólares para la prueba de sangre oculta en heces, hasta 1 000 dólares para la colonoscopia. Sin embargo, el costo por año de vida que se economiza con el tamizaje no difiere demasiado en todas las pruebas; todas se encuentran dentro de un margen aceptable de los estándares de EE.UU. en 2011 (21). Esto es así por varias razones. En primer lugar, las pruebas más económicas y sencillas se han de realizar más a menudo. Las pruebas de sangre oculta en heces se recomiendan que se efectúen anualmente; en cambio, la colonoscopia se recomienda cada 10 años. Con las pruebas más económicas también se obtienen más resultados falsos positivos que hacen que se realicen más pruebas (y, por lo tanto, más costos). Por último, se obtienen más resultados falsos negativos, y se omiten pacientes que realmente tienen carcinomas, lo cual comporta un costo para la atención de pacientes con carcinomas más avanzados.

Seguridad

Es ético y razonable aceptar que puede haber un cierto riesgo al realizar las pruebas de diagnóstico en enfermos que buscan ayuda para problemas específicos. Sin embargo, el médico no puede posponer la acción cuando el paciente está gravemente enfermo, y hace todo cuanto puede. Someter a

booksmedicos.org

riesgos a los pacientes que se supone que están bien es bastante diferente, porque en estas circunstancias el procedimiento debe ser extremadamente seguro. Esto se debe, en parte, a que las posibilidades de detectar la enfermedad en personas sanas son muy escasas. Así, aunque la colonoscopia apenas se considera un procedimiento peligroso cuando se utiliza en pacientes con problemas gastrointestinales, puede causar una perforación intestinal. De hecho, cuando la colonoscopia con un índice de perforaciones de 2 por 1 000 se utiliza para detectar el cáncer colorrectal en personas de 50 años, se producen casi dos perforaciones por cada cáncer detectado.

Han surgido inquietudes sobre los posibles riesgos a largo plazo con el mayor uso de la TC para el tamizaje de la enfermedad arterial coronaria o, en el caso de las exploraciones de todo el cuerpo, para una variedad de anomalías. La dosis de radiación de las TC varían según el tipo, y una TC para detectar el calcio coronario en promedio equivale a unas 30 radiografías de tórax, mientras que una exploración de cuerpo completo equivale a cerca de unas 120. Una estimación del riesgo proyecta un excedente de 29 000 carcinomas como resultado de los 70 millones de TC realizadas en EE.UU. en el 2007 (22). Si estas inquietudes son correctas, las TC utilizadas para el tamizaje del cáncer temprano podrían por sí mismas provocar carcinoma en las décadas siguientes.

Aceptable para médicos y pacientes

Si una prueba de tamizaje conlleva molestias, se suele tardar varios años en convencer a un gran porcentaje de pacientes para realizarla. Así ha sucedido con las citologías exfoliativas, las mamografías, las sigmoidoscopias y las colonoscopias. Sin embargo, en general, el público estadounidense apoya el tamizaje.

La aceptabilidad de la prueba puede ser un criterio que los médicos que la realizan normalmente pasan por alto. La aceptación médica es especialmente relevante en las pruebas de tamizaje especializadas, como las mamografías, las sigmoidoscopias o las colonoscopias. En una encuesta de 53 centros de mamografía, el 44 % indicó escasez de personal para realizar la mamografía. Los autores especularon que las razones pueden ser el menor reembolso para las mamografías de tamizaje, el aumento de los juicios por mala praxis en la radiología mamaria y las regulaciones administrativas (23).

booksmedicos.org

POSIBLES EFECTOS ADVERSOS DEL TAMIZAJE

Los efectos adversos de las pruebas de tamizaje son: malestar durante el procedimiento de la prueba en sí (la mayoría de mujeres que se someten a una mamografía dicen que el procedimiento es doloroso, aunque normalmente no tanto como para que las pacientes rechacen la prueba), efectos de la radiación a largo plazo tras la exposición a los procedimientos radiológicos, resultados falsos positivos de las pruebas (con los consiguientes abordajes diagnósticos innecesarios y los efectos negativos del etiquetado), sobrediagnóstico e incidentalomas. Estos dos últimos se estudian con más detalle en la siguiente sección.

Riesgo de un resultado falso positivo

Un **resultado falso positivo en la prueba de tamizaje** es un resultado anómalo en una persona no enferma. Como se ha señalado previamente, es probable que las pruebas con valores predictivos bajos (consecuencia de una baja prevalencia de la enfermedad, una reducida especificidad de la prueba o ambas) causen una mayor frecuencia de falsos positivos. Estos, a su vez, pueden provocar efectos de etiquetado negativos, molestias y gastos para obtener procedimientos de seguimiento. En ciertas situaciones, los resultados falsos positivos pueden conducir a una cirugía mayor. En un estudio sobre el tamizaje del cáncer de ovario, el 8,4 % (3 285) de 39 000 mujeres tuvieron un resultado falso positivo y un tercio de ellas se sometieron a cirugía como parte de la evaluación diagnóstica del resultado de la prueba. A causa de los falsos positivos en las pruebas de tamizaje, cinco veces más mujeres sin cáncer de ovario se sometieron a cirugía que aquellas con cáncer de ovario (24).

Tabla 10-4

Relación entre el número de diferentes pruebas de tamizaje solicitadas y el porcentaje de personas normales con al menos un resultado de la prueba anómalo

Número de pruebas	Personas con al menos una anormalidad (%)
1	5

booksmedicos.org

5	23
20	64
100	99,4

Datos de Sackett DL. Clinical diagnosis and the clinical laboratory. Clin Invest Med 1978;1:37-43.

Los resultados falsos positivos sólo son una pequeña minoría de los resultados de las pruebas de tamizaje (únicamente un 10 % de las mamografías dan resultados falsos positivos). Incluso así, pueden afectar a un gran porcentaje de personas, a las que se les ha efectuado la prueba de tamizaje, de dos maneras diferentes. La mayoría de los médicos no realizan sólo una o dos pruebas a los pacientes que acuden para hacerse una revisión rutinaria. La tecnología moderna, y quizá el temor a las demandas, ha fomentado la tendencia de «cubrir todas las bases». Los análisis de sangre automatizados facilitan a los médicos la solicitud de varias pruebas con sólo hacer una señal en el recuadro adecuado.

Cuando los resultados de las pruebas de tamizaje se expresan en escalas de intervalo (como en la mayoría de los análisis de sangre), y cuando la *normalidad* se define por el intervalo cubierto por el 95 % de los resultados (como es habitual), cuantas más pruebas solicite el médico, mayor será el riesgo de obtener resultados falsos positivos. De hecho, como se muestra en la tabla 10-4, si el médico solicita suficientes pruebas, se detectarán «anomalías» en prácticamente todos los pacientes sanos. (Se publicó una parodia sobre este fenómeno, titulada *The Last Well Person*) (25).

Otra razón por la que muchas personas pueden obtener un resultado falso positivo de una prueba de tamizaje es que la mayoría de estas pruebas se repiten con regularidad. En cada tamizaje, el paciente se expone a tener un resultado falso positivo.

Ejemplo

En un ensayo clínico sobre el tamizaje del cáncer de pulmón con una TC espiral en dosis bajas o una radiografía de tórax, la primera ronda de tamizaje causó resultados falsos positivos en el 21 % de las personas sometidas al tamizaje con TC y el 9 % de las sometidas al tamizaje con radiografía; después de la segunda ronda, la proporción total de personas que experimentó un

booksmedicos.org

resultado positivo aumentó al 33 % en el grupo de TC y un 15 % en el grupo de radiografía de tórax (26). En un estudio relacionado, los participantes se sometieron a pruebas de tamizaje para el cáncer de próstata, ovario y colorrectal, así como radiografías de tórax para cáncer de pulmón, para un total de 14 pruebas a lo largo de 3 años. El riesgo acumulativo de tener al menos un resultado falso positivo en la prueba de tamizaje fue del 60,4 % para los hombres y del 48,8 % para las mujeres. El riesgo de ser sometido a un procedimiento diagnóstico invasivo debido a un resultado falso positivo fue del 28,5 % para los hombres y del 22,1 % para las mujeres (27).

Riesgo negativo de etiquetado

Los resultados de las pruebas a veces pueden tener efectos psicológicos importantes en los pacientes, denominado **efecto de etiquetado.** Un buen resultado en las pruebas de tamizaje no comporta ningún efecto de etiquetado, ni positivo ni negativo.

Puede producirse un efecto de etiquetado positivo cuando a un paciente se le comunica que todos los resultados de la prueba de tamizaje son normales. Muchos facultativos han recibido respuestas parecidas a esta: «Perfecto, esto quiere decir que puedo trabajar un año más.» Por otro lado, cuando a un paciente se le comunica que el resultado de la prueba de tamizaje presenta anomalías y es necesario realizar más pruebas, el efecto psicológico puede ser adverso, en particular la detección de cáncer. Algunas personas con pruebas falsas positivas continúan preocupadas incluso después que se les ha dicho que todo era normal en las pruebas de seguimiento. Dado que este es un grupo de personas *no* enfermas, desde el punto de vista ético los efectos de etiquetado negativos son muy preocupantes. En esta situación, los resultados del tamizaje pueden potenciar la vulnerabilidad en lugar de la salud, y pueden perjudicar en lugar de beneficiar.

Ejemplo

Un estudio sobre los hombres con resultados anormales en la prueba de tamizaje del PSA, que posteriormente se declararon libres del cáncer tras los estudios diagnósticos, encontró que 1 año después del seguimiento, el 26 % expresó preocupación sobre el cáncer de próstata, en comparación con el 6 % de los hombres con resultados normales en la prueba de PSA, y el 46 % informaron que sus esposas o sus allegados estaban preocupados sobre el cáncer de próstata en comparación con el 14 % de los que tuvieron resultados normales (28). La ansiedad también se produce con la obtención de resultados de tamizaje anómalos para otros cánceres, así como

booksmedicos.org

con la determinación de la presión arterial. En resumen, las personas no responden bien cuando se les informa que el resultado de su prueba de tamizaje no fue normal y deben realizarse más pruebas.

Los efectos del etiquetado a veces son imprevisibles, especialmente entre las personas que saben que tienen alto riesgo de enfermedades genéticas debido a su historia familiar. Los estudios de familiares de pacientes con enfermedad de Huntington, trastorno neurológico que se manifiesta en la edad madura y causa deterioro mental que conduce a la demencia, trastornos del movimiento y la muerte, han encontrado que la salud psicológica después de las pruebas genéticas no se deteriora en aquellos cuyas pruebas son positivas, quizá porque ya habían salido de la incertidumbre. Los estudios de mujeres a las que se les realizan pruebas de mutaciones genéticas que aumentan el riesgo de sufrir carcinomas de mama y ovario han revelado que en las mujeres que son positivas a la mutación experimentan un poco o ningún deterioro psicológico.

Riesgo de seudoenfermedad (sobrediagnóstico) en el tamizaje del cáncer

La justificación del tamizaje (detección sistemática) del cáncer es que cuanto antes se detecte, más posibilidades hay de que se cure. Por lo tanto, siempre es mejor detectar el cáncer lo antes posible. Esta tesis está siendo cuestionada por la observación de que la incidencia aumenta a menudo después de introducir el tamizaje generalizado para un cáncer en particular. Es previsible observar un aumento temporal de la incidencia porque la prueba de detección adelanta el tiempo de diagnóstico, añadiendo casos tempranos al número habitual de carcinomas prevalentes que se diagnostican sin tamizaje, pero el impacto temporal de la incidencia debe disminuir hasta el nivel inicial después de unos cuantos años. Sin embargo, con varios tipos de carcinomas, la incidencia ha permanecido en un nivel elevado, como se muestra en la figura 2-5 para el cáncer de próstata. Es como si el tamizaje causara más carcinomas.

Algunos cánceres crecen con lentitud, y otros incluso remiten, por lo que no causan ningún problema al paciente. Si estos cánceres se detectan con el tamizaje, se denominan seudoenfermedad; el proceso que conduce a su detección se denomina **sobrediagnóstico,** porque el hecho de localizarlos no ayuda al paciente. El sobrediagnóstico es un ejemplo

booksmedicos.org

extremo de sesgo de anticipación diagnóstica (fig. 10-8). Los cánceres encontrados tienen un pronóstico tan bueno que nunca serían evidentes sin la tecnología de tamizaje. Algunas estimaciones son que el 50 % de cánceres de próstata diagnosticados por tamizaje son debidos a sobrediagnóstico.

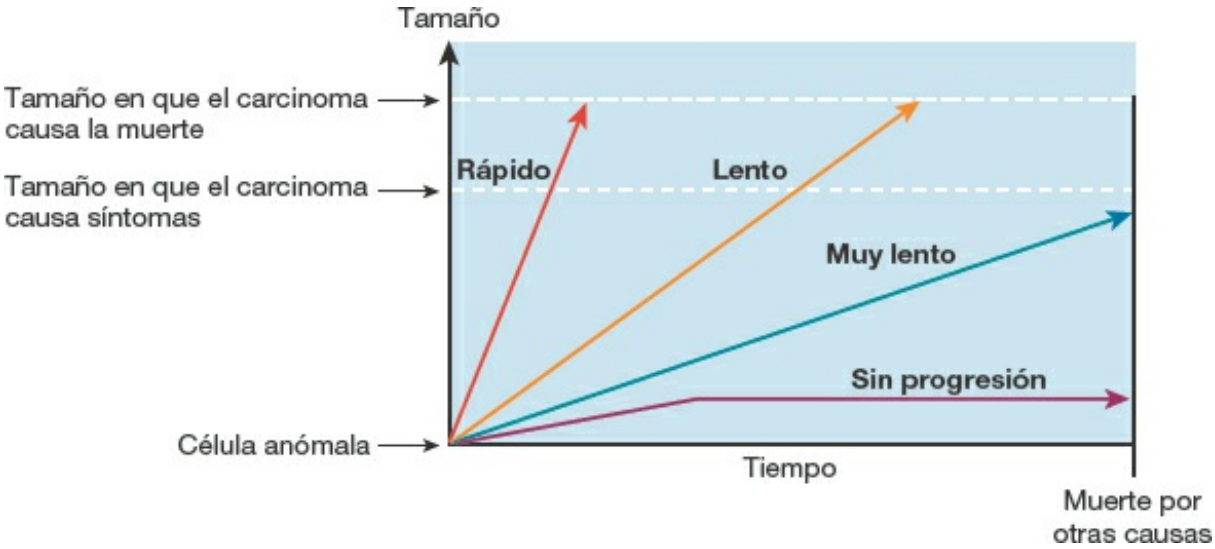


Figura 10-8 ■ Mecanismo de sobrediagnóstico en el tamizaje de carcinomas. Obsérvese que los carcinomas que no progresan, así como algunos de crecimiento muy lento, nunca causan problemas clínicos. Cuando estos carcinomas se detectan mediante el tamizaje, se produce el sobrediagnóstico, que es una forma extrema de sesgo de anticipación diagnóstica. (Redibujado con autorización de Welsh HG. Should I Be Tested for Cancer? Maybe Not and Here's Why. Berkeley, CA: University of California Press; 2004.)

A medida que la investigación genética aclara el desarrollo de carcinomas, parece que una secuencia de cambios genéticos acompaña a la patogenia desde el tejido normal a la enfermedad maligna. En cada etapa, sólo algunas lesiones pasan a la siguiente fase de carcinogénesis. Es probable que se produzca sobrediagnóstico porque las pruebas de tamizaje más modernas detectan las lesiones al inicio de la cadena. El objetivo consiste en diferenciar los cánceres iniciales que provocarán mortalidad y morbilidad de las que permanecerán latentes a lo largo de toda la vida, aunque patológicamente parecen ser lo mismo. Actualmente, la tecnología de tamizaje no es capaz de hacer esto.

Para determinar si se produce el sobrediagnóstico y en qué grado, es necesario precisar los índices de mortalidad específicos de la enfermedad (no índices de supervivencia). Esto puede hacerse mediante ensayos

booksmedicos.org

aleatorizados a largo plazo o con estudios observacionales cuidadosos basados en la población.

Ejemplo

El neuroblastoma, un tumor del tejido nervioso próximo al riñón, es el segundo tumor infantil más frecuente. El pronóstico depende del estado de la enfermedad y es mejor si el tumor se diagnostica en la infancia. El tratamiento consiste en cirugía y quimioterapia. Para detectar el tumor, se puede utilizar un simple análisis de orina que determine catecolaminas y metabolitos. Varios estudios japoneses demostraron una mejora de los índices de supervivencia tras el tamizaje, pero el tiempo de anticipación diagnóstica y un grupo de control histórico podrían haber sesgado los resultados. Además, no había ningún registro basado en la población de carcinomas infantiles en Japón que garantizara la verificación de todos los casos de neuroblastoma. Por último, algunos casos de esta afección remiten sin tratamiento, lo que planteó la posibilidad del sobrediagnóstico.

En Alemania (29) y en Quebec (30) se realizaron dos estudios basados en la población, en los que se ofreció el tamizaje a todos los niños de algunas zonas, y a los de otras no, de modo que actuaban como controles concurrentes. Se utilizaron los registros de tumores existentes para hacer un seguimiento de todos los casos y de las muertes debidas al neuroblastoma. En ambos estudios, la incidencia de neuroblastoma fue el doble en los grupos de tamizaje, pero las tasas de mortalidad por neuroblastoma en los siguientes 5 años fueron equivalentes en los grupos sometidos al tamizaje y los que no se sometieron. Parece que la prueba de detección, ante todo, encontró los tumores con un pronóstico favorable, muchos de los cuales habrían remitido si no se hubieran detectado. Mientras tanto, no se percibía la enfermedad sumamente invasiva. Los investigadores de los dos estudios concluyeron que el tamizaje infantil para detectar neuroblastomas conlleva el sobrediagnóstico, pero no reduce la mortalidad por esta afección.

Se ha demostrado sobrediagnóstico en los ensayos aleatorizados para el tamizaje del cáncer de pulmón y mama. También puede ocurrir cuando se detectan anomalías premalignas en el tamizaje del cáncer de cuello uterino, colorrectal y de mama, es decir, la displasia cervical, los pólipos adenomatosos y el carcinoma ductal *in situ* (anomalías que a veces se denominan **enfermedad preexistente).** Es importante comprender que el sobrediagnóstico puede coexistir con el tamizaje efectivo y que, aunque los ensayos aleatorizados y los estudios poblacionales pueden ayudar a determinar la cantidad de sobrediagnóstico, es imposible identificarlo en un paciente individual.

Incidentalomas

booksmedicos.org

Durante las últimas dos décadas se ha vuelto más habitual el uso de la TC como prueba de tamizaje. Esta prueba se ha evaluado rigurosamente como una «colonoscopia virtual» para el tamizaje del cáncer colorrectal y también para el tamizaje del cáncer de pulmón. Se ha recomendado como una prueba de detección para la enfermedad coronaria (con las puntuaciones de calcio) y para el tamizaje general con la TC de cuerpo completo. A diferencia de la mayoría de las pruebas de tamizaje, la TC a menudo visualiza mucho más que el área determinada de interés. Por ejemplo, la colonoscopia por TC muestra el abdomen y la parte inferior del tórax. En el proceso, a veces se detectan anomalías fuera del colon. Las masas o las lesiones detectadas incidentalmente por una exploración radiológica se denominan **incidentalomas.**

Ejemplo

Una revisión sistemática de 17 estudios encontró que los incidentalomas eran habituales en la colonoscopia por TC; el 40 % de los 3 488 pacientes tuvieron al menos un hallazgo incidental, y el 14 % tuvieron una evaluación posterior (31). Los estudios diagnósticos descubrieron cánceres no colónicos en etapas tempranas en aproximadamente el 1 % de los pacientes y aneurismas de la aorta abdominal con un tamaño > 5,5 cm en un 0,1 %. Por lo tanto, los incidentalomas fueron muy habituales, pero comprendían algunos trastornos potencialmente importantes. No hubo información sobre el número de pacientes que no se sometieron a otra evaluación y después sufrieron consecuencias serias. Además, la mayoría de los carcinomas detectados fueron de un tipo no recomendado para el tamizaje. Con evidencia insuficiente sobre los beneficios y la probabilidad del aumento de los costos en la búsqueda de resultados extracolónicos, el Center for Medicare y Medicaid Services se negaron a cubrir la colonoscopia con TC para el tamizaje en la población de Medicare.

CAMBIOS EN LAS PRUEBAS DE TAMIZAJE Y TRATAMIENTOS A LO LARGO DEL TIEMPO

La evaluación cuidadosa del tamizaje ha sido particularmente difícil cuando las pruebas aprobadas para fines diagnósticos se utilizan como pruebas de tamizaje, antes de ser evaluadas en un estudio adecuado, un problema similar al uso de las intervenciones terapéuticas para la prevención sin haberlas evaluado para este propósito. Por ejemplo, las pruebas de TC y la resonancia magnética se crearon con fines diagnósticos para pacientes con molestias graves o enfermedades conocidas, y el PSA se

booksmedicos.org

desarrolló para determinar si el tratamiento para el cáncer de próstata había sido un éxito. Todas estas pruebas ahora se usan habitualmente como pruebas de tamizaje, pero la mayoría se volvieron habituales en la práctica sin una evaluación cuidadosa. Solo las TC de dosis baja para el tamizaje del cáncer de pulmón fue sometida a una evaluación cuidadosa antes de generalizar su uso. El tamizaje con PSA se volvió tan popular en EE.UU. que, cuando fue sujeto de un ensayo aleatorizado cuidadoso, más de la mitad de los hombres asignados al grupo control se realizaron una prueba de PSA durante la duración del ensayo. Cuando las pruebas se utilizan de forma tan habitual, es difícil determinar si son efectivas de manera rigurosa.

Con el tiempo, las mejorías en las pruebas de tamizaje, los tratamientos y las vacunas pueden cambiar la necesidad de tamizaje. Como se comentó previamente, la prevención secundaria efectiva es un proceso de dos pasos; una buena prueba de tamizaje seguida de un buen tratamiento para aquellos en quienes se encontró la enfermedad. Los cambios en cualquiera de ellos pueden afectar a la eficacia del tamizaje para prevenir la enfermedad. En un extremo, una prueba de tamizaje de gran precisión no ayudará a prevenir los resultados adversos de la enfermedad si no hay un tratamiento efectivo. Las pruebas de tamizaje para el VIH antecedieron al desarrollo de un tratamiento efectivo para el VIH; por lo tanto, a principios de la historia del VIH, el tamizaje no podía prevenir la progresión de la enfermedad en los afectados. Con el desarrollo de tratamientos cada vez más efectivos, ha aumentado el tamizaje para el VIH. En el otro extremo, un tratamiento altamente efectivo puede hacer que el tamizaje sea innecesario. Con el tratamiento moderno, la tasa de supervivencia a 10 años para el carcinoma testicular es de un 85 %, tan alto que sería difícil mostrar una mejoría con el tamizaje de este tipo de cáncer poco habitual. Conforme la vacunación contra el VPH para la prevención del cáncer de cuello uterino sea más generalizada y capaz de cubrir todos los tipos carcinogénicos de VPH, la necesidad de tamizaje para este cáncer debería disminuir con el tiempo. Algunos estudios recientes sobre la mamografía de tamizaje no han encontrado los beneficios esperados de la mortalidad que se observaron en estudios anteriores, en parte debido a que la mortalidad por cáncer de mama entre las mujeres que no se sometieron al tamizaje fue menor que en el pasado, probablemente por los mejores tratamientos. Por lo tanto, con la introducción de nuevas terapias y pruebas de tamizaje, la efectividad del

booksmedicos.org

tamizaje cambiará y será necesaria una reevaluación continua.

PONDERAR LOS BENEFICIOS CONTRA LOS PERJUICIOS DE LA PREVENCIÓN

¿Cómo se pueden combinar los muchos aspectos de la prevención comentados en este capítulo para tomar una decisión sobre si se debe incluir una intervención preventiva en la práctica clínica? Conceptualmente, la decisión debe basarse en sopesar la *magnitud de los beneficios contra la magnitud de los perjuicios* que ocurrirán como resultado de una acción. Este abordaje se ha vuelto habitual al tomar decisiones sobre el tratamiento; los informes sobre los ensayos aleatorizados suelen incluirse en los perjuicios y los beneficios.

Un enfoque directo es presentar los beneficios y los perjuicios de una actividad preventiva particular de manera ordenada y comprensible. Siempre que sea posible, se deben presentar utilizando riesgos absolutos, no relativos. La figura 10-9 resume los principales beneficios y perjuicios estimados de la mamografía anual para las mujeres en la década de los 40, 50 y 60 años (7,32,33). Este enfoque puede ayudar a los médicos y a los pacientes a comprender lo que conlleva cuando se toma la decisión de realizar el tamizaje. También puede ayudar a aclarar por qué los diferentes individuos y grupos de expertos toman diferentes decisiones sobre una actividad preventiva, incluso cuando revisan el mismo grupo de información. Diferentes personas asignan valores diferentes a los beneficios y a los perjuicios (34).

Otro enfoque para considerar los beneficios y los perjuicios es un proceso de modelado que los expresa en una sola medición y luego resta los perjuicios de los beneficios. (La medición más usada son los **años de vida ajustados por calidad [AVAC]).** La ventaja de este enfoque es que se pueden comparar los tipos diferentes de prevención (p. ej., vacunas, tamizaje del cáncer colorrectal y tratamientos terciarios para la diabetes) entre sí, que es importante para las autoridades con recursos limitados. La desventaja es que para la mayoría de los médicos y las autoridades es difícil comprender el proceso mediante el cual se tratan los beneficios y los perjuicios.

Sin importar el método utilizado para ponderar los beneficios y los perjuicios de las actividades preventivas, se debe evaluar la calidad de la

booksmedicos.org

evidencia para cada beneficio y perjuicio con el fin de evitar el problema de la «basura que entra y la basura que sale». Varios grupos que elaboran recomendaciones para la prevención clínica han desarrollado métodos explícitos para evaluar la evidencia y considerar la fuerza de la evidencia al hacer sus recomendaciones (v. cap. 14).

Si los beneficios de una actividad preventiva superan a los perjuicios, el paso final es determinar el efecto económico de aplicarla. A algunos analistas les gusta afirmar que la «prevención ahorra dinero», pero sólo lo hace en raras ocasiones. (Una posible excepción es el tamizaje del cáncer colorrectal. La quimioterapia para esta enfermedad se ha vuelto tan costosa que algunos análisis ahora encuentran que el tamizaje para este cáncer ahorra dinero). Aun así, la mayoría de los servicios preventivos recomendados por los grupos que han evaluado cuidadosamente los datos tienen el mismo costo-efectividad que otras actividades clínicas.

El **análisis de costo-efectividad** es un método para evaluar los costos y los beneficios para la salud de una intervención. Deben considerarse todos los costos relacionados con el tratamiento y la aparición de la enfermedad, tanto con y sin la actividad preventiva, así como todos los costos relacionados con la actividad preventiva en sí misma. Después se calculan los beneficios para la salud de la actividad y se determina el incremento del costo para cada unidad de beneficio.

Ejemplo

El cáncer de cuello uterino está causado por la infección persistente de las células epiteliales por ciertos genotipos de VPH. Se han desarrollado vacunas contra los genotipos carcinogénicos de VPH que causan un 70 % de todos los cánceres de cuello uterino. Un ensayo aleatorizado detectó que la vacuna era altamente efectiva para prevenir los precursores y el cáncer temprano de cuello uterino en mujeres sin evidencia previa de infección por VPH. Se realizó un análisis de costo-efectividad para evaluar los beneficios y los costos de vacunar a las niñas de 12 años de edad que luego se someterían a las pruebas para el cáncer de cuello uterino en la edad adulta (35). Los investigadores estimaron los costos médicos desde una perspectiva social que incluían los costos de la atención médica (p. ej., vacunación y tamizaje, así como los costos del diagnóstico y el tratamiento para el cáncer de cuello uterino) y los indirectos (p. ej., tiempo laboral que se tomaron los pacientes para las consultas médicas). El beneficio se expresó en términos de AVAC ganado por la combinación de la vacuna más el tamizaje en comparación con la falta de tamizaje o vacunación. En el análisis se modificó la edad y los intervalos en los que tendría lugar el tamizaje, así como la eficacia de la vacuna a lo largo del tiempo. Considerando una eficacia de la vacuna del 90 %, el análisis estimó que el inicio del tamizaje

booksmedicos.org

convencional con la citología de Papanicolaou a los 25 años y la repetición del tamizaje en un intervalo de 3 años, podrían reducir el riesgo de cáncer de cuello uterino a lo largo de la vida en un 94 %, con un costo de 58 500 dólares por AVAC. Si el tamizaje comenzaba a los 21 años de edad y se repetía cada 5 años, se podía reducir el riesgo de cáncer a lo largo de la vida en un 90 %, con un costo de 57 400 dólares por AVAC. En resumen, los resultados sugieren que la combinación de la vacuna de VPH y el tamizaje para el cáncer de cuello uterino pueden disminuir el riesgo de cáncer, son costo-efectivos y permiten retrasar el inicio del tamizaje y/o disminuir su frecuencia.

booksmedicos.org

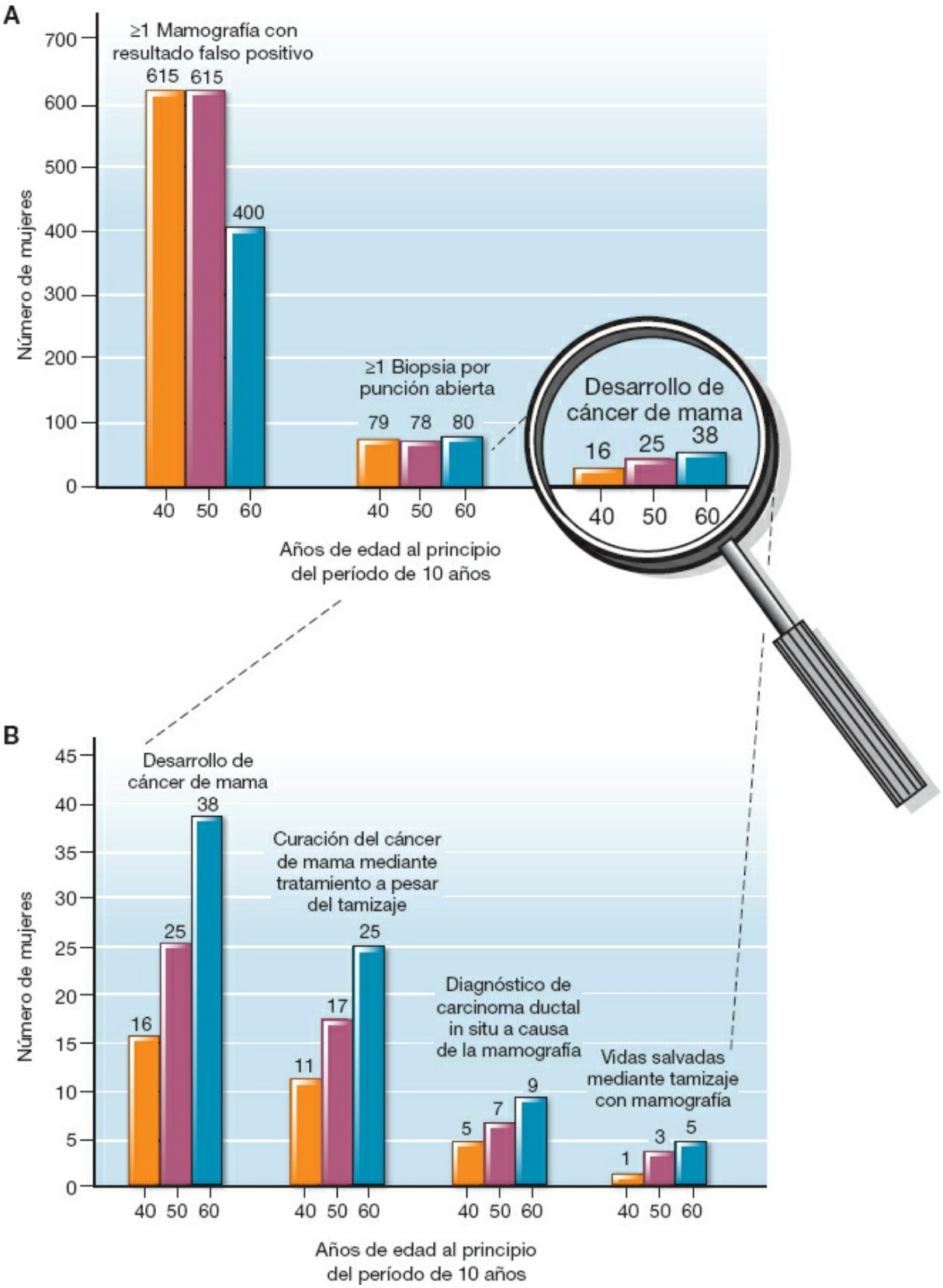


Figura 10-9 ■ Ponderar los beneficios y los perjuicios en el momento de decidir sobre una

booksmedicos.org

actividad preventiva: comparar los beneficios y los perjuicios estimados de la mamografía de tamizaje. Posibilidades entre 1 000 mujeres de 40, 50 o 60 años que se someten a una mamografía anual durante 10 años: **(A)** mamografía con resultado falso positivo, biopsia mamaria y desarrollo de cáncer de mama, y **(B)** curación del cáncer de mama a pesar de la prueba de detección, habiéndose diagnosticado un carcinoma ductal *in situ* no invasivo y evitándose la muerte por carcinoma mamario debido a la mamografía de tamizaje. (Los estimados se calculan a partir de las referencias 7, 32 y 33.)

El esfuerzo para reunir toda la información necesaria con el fin de tomar una decisión para realizar una actividad preventiva en la práctica clínica no es algo que un solo médico puede lograr, pero al revisar las recomendaciones sobre la prevención, los médicos individuales pueden determinar si los beneficios y los perjuicios de una actividad se presentan de manera comprensible y si la recomendación ha considerado la fortaleza de la evidencia. También pueden considerar las estimaciones de costo-efectividad. Con estas evidencias, deben ser capaces de compartir con sus pacientes la información que necesitan. Los pacientes pueden entonces tomar una decisión informada acerca de las actividades preventivas que tienen en cuenta la información científica y sus valores individuales.

Preguntas de revisión

Lea los escenarios relacionados y seleccione la mejor respuesta para las preguntas 10-1 a 10-6.

Se realizó un estudio para determinar si la prueba de tamizaje de sangre oculta en heces reducía la mortalidad del cáncer colorrectal (11). Se asignaron personas de entre 50 y 60 años de edad, de forma aleatoria, a la prueba de detección o al grupo de control, realizándose un seguimiento durante 13 años. En este período, entre las 15 570 personas a las que se les asignó de forma aleatorizada la prueba de detección anual, hubo 323 casos de cáncer y 82 muertes por cáncer colorrectal, y en el grupo de control de 15 394 personas, hubo 356 casos de cáncer y 121 muertes por cáncer colorrectal. Las investigaciones de las pruebas con resultados positivos revelaron que un 30 % del grupo tamizaje tenía pólipos en el colon. La sensibilidad y la especificidad de la prueba para el cáncer de colon fue en ambos casos del 90 %.

booksmedicos.org

10-1. ¿Cuál es la reducción del riesgo relativo de la mortalidad por cáncer colorrectal en el grupo de tamizaje?

A. 33 %.
B. 39 %.
C. 48 %.

10-2. ¿Cuántos pacientes necesitarían someterse a las pruebas de tamizaje en los siguientes 13 años para prevenir una muerte por cáncer colorrectal?

A. 43.
B. 194.
C. 385.

10-3. El hecho de que el 30 % del grupo tamizaje tenga pólipos colónicos sugiere todo lo siguiente, *excepto:*

A. Se investigó al menos al 30 % del grupo tamizaje debido a las pruebas positivas de sangre oculta en heces.
B. El valor predictivo positivo de la prueba fue bajo.
C. El valor predictivo negativo de la prueba fue bajo.

En un estudio controlado y aleatorizado de tamizaje mediante radiografía de tórax y estudio citológico de esputo para cáncer broncopulmonar, se asignaron de forma aleatoria 9 000 hombres al grupo de tamizaje durante 6 años o al grupo de control (16). Al cabo de 20 años, la mortalidad por carcinomas broncopulmonares era la misma en ambos grupos (4,4/1 000 personas-años en el grupo tamizaje y 3,9/1 000 personas-años en el grupo de control). Sin embargo, la supervivencia media de los pacientes diagnosticados con cáncer de pulmón fue de 1,3 años en el grupo de tamizaje y de 0,9 años en el grupo control. Asimismo, el tamizaje detectó más casos de carcinomas broncopulmonares: en el grupo de tamizaje se diagnosticaron 206 y en el grupo de control, 160.

10-4. ¿Cuál es la mejor conclusión después de leer este estudio?

A. No tiene sentido encontrar una mejor tasa de supervivencia sin cambios en la tasa de mortalidad por cáncer de pulmón y el

booksmedicos.org

estudio no es válido.

B. Debido a que la mortalidad no ha cambiado, el tamizaje puede haber dado lugar a un mayor «tiempo de enfermedad» para las personas diagnosticadas con cáncer de pulmón.

C. La mejoría en la supervivencia demuestra que el tamizaje fue efectivo en el estudio.

10-5. ¿Qué sesgo explicaría con mayor precisión la mejoría en la supervivencia ante la ausencia de mejoría en la tasa de mortalidad en este estudio?

A. Sesgo de anticipación diagnóstica.

B. Sesgo de supervivencia.

C. Sesgo de cumplimiento.

D. Sesgo de duración.

10-6. ¿Cuál es la razón más probable para haber encontrado 206 carcinomas de pulmón en el grupo de tamizaje y solamente 106 en el grupo control?

A. Había más fumadores en el grupo de tamizaje.

B. El tamizaje encontró más carcinomas tempranos y el número de carcinomas en el grupo control se recuperará con el tiempo.

C. El tamizaje encontró algunos carcinomas que no habrían recibido atención médica sin el tamizaje.

Seleccione la mejor respuesta para las preguntas 10-7 a 10-11.

10-7. Cuando se evalúa una nueva vacuna, ¿cuál de los siguientes factores es menos importante?

A. La eficacia para prevenir la enfermedad.

B. La seguridad de la vacuna.

C. El peligro de la enfermedad.

D. El costo de administrar la vacuna.

10-8. Cuando se usa la misma prueba para fines diagnósticos y de tamizaje, ¿cuál de las siguientes afirmaciones es la más correcta?

booksmedicos.org

A. Es probable que la sensibilidad y la especificidad sean iguales en ambas situaciones.

B. El valor predictivo positivo será mayor para el tamizaje.

C. La prevalencia de la enfermedad será más alta en el diagnóstico.

D. El sobrediagnóstico es igualmente probable en ambas situaciones.

10-9. Un estudio encontró que los voluntarios para someterse a una nueva prueba de tamizaje tenían mejores resultados de salud que las personas que se negaron a someterse a la prueba. ¿Cuál de las siguientes afirmaciones explica con mayor probabilidad este hallazgo?

A. Las personas que no aceptan el tamizaje generalmente están más sanas que aquellos que aceptan.

B. Es más probable que los voluntarios para el tamizaje necesiten más la prueba que aquellos que se niegan.

C. Los voluntarios tienden a estar más interesados en su salud que aquellos que no participan en las actividades preventivas.

10-10. Las siguientes afirmaciones son correctas, *excepto:*

A. El patrón de referencia de una prueba con fines diagnósticos puede ser diferente para la misma prueba cuando se usa con fines de tamizaje.

B. El método de incidencia no se puede utilizar para calcular la sensibilidad de las pruebas de tamizaje del cáncer.

C. Cuando se inicia un programa de tamizaje, la mayoría de las personas con la enfermedad se detectan en la primera ronda de tamizaje que en rondas posteriores.

10-11. Para un análisis de costo-efectividad de una actividad preventiva, ¿qué tipos de costos deben incluirse?

A. Costos médicos, como aquellos asociados a administrar la intervención preventiva.

B. Todos los costos médicos, incluyendo el seguimiento de las pruebas positivas y el tratamiento de las personas diagnosticadas con la enfermedad, con y sin la actividad preventiva.

booksmedicos.org

C. Los costos indirectos, como la pérdida de ingresos debido a las faltas laborales, entre los pacientes que reciben la prevención y aquellos que desarrollan la enfermedad.

D. Los costos indirectos tanto para los pacientes como para los cuidadores.

E. Todos los anteriores.

Encontrará las respuestas en el apéndice A.

BIBLIOGRAFÍA

1. Schappert SM, Rechtsteiner EA. Ambulatory medical care utilization estimates for 2007. National Center for Health Statistics. *Vital Health Stat* 2011;13(169). Available at http://www.cdc.gov/nchs/data/series/sr_13/sr13_169.pdf. Accessed January 11, 2012.
2. Prevention. 2011. In Meriam-webster.com. Available at http://www.merriam-webster.com/dictionary/prevention. Accessed January 13, 2012.
3. Siegel R, Ward E, Brawley O, et al. Cancer statistics, 2011. The impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J Clin* 2011;61: 212–236.
4. National Center for Health Statistics. Health. United States. 2010; With special feature on death and dying. Hyattsville, MD. 2011.
5. Roe MT, Ohman EM. A new era in secondary prevention after acute coronary syndrome. *N Engl J Med* 2012;366:85–87.
6. Harris R, Sawaya GF, Moyer VA, et al. Reconsidering the criteria for evaluating proposed screening programs: reflections from 4 current and former members of the U.S. Preventive Services Task Force. *Epidemiol Rev* 2011;33:20–35.
7. Howlader N, Noone AM, Krapcho M, et al. (eds). SEER Cancer Statistics Review, 1975-2008, National Cancer Institute. Bethesda, MD. Available at http://seer.cancer.gov/csr/1975_2008/, based on November 2010 SEER data submission, posted to the SEER Web site, 2011. Accessed January 13, 2012.
8. Chang MH, You SL, Chen CJ, et al. Decreased incidence of hepatocellular carcinoma in hepatitis B vaccinees: a 20 year follow-up study. *J Natl Cancer Inst* 2009;101:1348–1355.
9. Greene SK, Rett M, Weintraub ES, et al. Risk of confirmed Guillain-Barré Syndrome following receipt of monovalent inactivated influenza A (H1N1) and seasonal influenza vaccines in the Vaccine Safety Datalink Project, 2009-2010. *Am J Epidemiol* 2012;175:1100–1109.
10. Fiore MC, Jaén CR, Baker TB, et al. Treating tobacco use and dependence: 2008 Update. Clinical Practice Guideline. Rockville, MD: *U.S. Department of Health and Human Services.* Public Health Service. May 2008.
11. Mandel JS, Bond JH, Church TR, et al. (for the Minnesota Colon Cancer Control Study). Reducing mortality from colorectal cancer by screening for fecal occult blood. *N Engl J Med* 1993;328:1365–1371.
12. Marcus PM, Bergstralh EJ, Fagerstrom RM, et al. Lung cancer mortality in the Mayo Lung Project: impact of extended follow-up. *J Natl Cancer Inst.* 2000;92:1308–1316.
13. The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.
14. Gøde P, Lund-Andersen H, Hans-Henrik P, et al. Effect of a multifactorial intervention on

booksmedicos.org

mortality in type 2 diabetes. *N Engl J Med* 2008;358:580–591.
15. Friedman GD, Collen MF, Fireman BH. Multiphasic health checkup evaluation: a 16-year follow-up. *J Chron Dis* 1986;39: 453–463.
16. Avins AL, Pressman A, Ackerson L, et al. Placebo adherence and its association with morbidity and mortality in the studies of left ventricular dysfunction. *J Gen Intern Med* 2010;25: 1275–1281.
17. Selby JV, Friedman GD, Quesenberry CP, et al. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *N Eng J Med* 1992;326:653–657.
18. Gann PH, Hennekens CH, Stampfer MJ. A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. *JAMA* 1995;273:289–294.
19. Delongchamps NB, Singh A, Haas GP. The role of prevalence in the diagnosis of prostate cancer. *Cancer Control* 2006;13: 158–168.
20. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med* 2003;138:168–175.
21. Lansdorp-Vogelaar I, Knudsen AB, Brenner H. Cost-effectiveness of colorectal cancer screening. *Epi Rev* 2011;33:88–100.
22. Berrington de González A, Mahesh M, Kim K-P, et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. *Arch Intern Med* 2009;169:2071–2077.
23. D'Orsi CD, Shin-Ping Tu, Nakano C. Current realities of delivering mammography services in the community: do challenges with staffing and scheduling exist? *Radiology* 2005;235: 391–395.
24. Buys SS, Partridge E, Black A, et al. Effect of screening on ovarian cancer mortality. The Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening randomized controlled trial. *JAMA* 2011;305:2295–2303.
25. Meador CK. The last well person. *N Engl Med J* 1994;330: 440–441.
26. Croswell JM, Baker SG, Marcus PM, et al. Cumulative incidence of false-positive test results in lung cancer screening: a randomized trial. *Ann Intern Med* 2010;152:505–512.
27. Croswell JM, Kramer BS, Kreimer AR. Cumulative incidence of false-positive results in repeated multimodal cancer screening. *Ann Fam Med* 2009;7:212–222.
28. Fowler FJ, Barry MJ, Walker-Corkery BS. The impact of a suspicious prostate biopsy on patients' psychological, socio-behavioral, and medical care outcomes. *J Gen Intern Med* 2006; 21: 715–721.
29. Schilling FH, Spix C, Berthold F, et al. Neuroblastoma screening at one year of age. *N Engl J Med* 2002;346:1047–1053.
30. Woods WG, Gao R, Shuster JJ, et al. Screening of infants and mortality due to neuroblastoma. *N Engl J Med* 2002;346: 1041–1046.
31. Xiong T, Richardson M, Woodroffe R, et al. Incidental lesions found on CT colonography: their nature and frequency. *Br J Radiol* 2005;78:22–29.
32. Hubbard RA, Kerlikowske K, Flowers CI, et al. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography. *Ann Intern Med* 2011;155:481–492.
33. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Ann Intern Med* 2009;151:738–747.
34. Gillman MW, Daniels SR. Is universal pediatric lipid screening justified? *JAMA* 2012;307:259–260.
35. Goldie SJ, Kohli M, Grima D. Projected clinical benefits and cost-effectiveness of a human

booksmedicos.org

papillomavirus 16/18 vaccine. *J Natl Cancer Inst* 2004;96:604–615.

booksmedicos.org

Capítulo 11

Azar

Es una práctica habitual juzgar que un resultado es significativo si su magnitud es tal que podría haberse producido por casualidad con una frecuencia no mayor de una vez en veinte ensayos. Este es un nivel de significación arbitrario pero conveniente para el investigador práctico, aunque no quiere decir que debe dejarse engañar una vez en cada veinte experimentos.

—Ronald Fisher
1929 (1)

PALABRAS CLAVE

Verificación de la hipótesis
Estimación
Estadísticamente significativo
Error de tipo I (o error α)
Error de tipo II (o error β)
Estadística deductiva o inferencial
Verificación estadística
Valor de *P*
Significación estadística
Pruebas estadísticas
Hipótesis nula
Pruebas no paramétricas
Prueba de dos colas
Prueba de una cola
Potencia estadística o poder estadístico
Tamaño muestral
Estimación puntual
Precisión estadística

booksmedicos.org

Intervalo de confianza
Comparaciones múltiples
Modelo multivariado
Razonamiento bayesiano

Hay dos procesos que dificultan el aprendizaje a partir de la experiencia clínica, bien durante la investigación formal, o en el curso de la asistencia al paciente: los sesgos y el azar. Como ya se comentó en el capítulo 1, el sesgo es un error sistemático, el resultado de cualquier proceso que provoca discrepancia sistemáticamente en las observaciones de los valores reales. Gran parte de este libro ha tratado de cuándo podría aparecer el sesgo, cómo evitarlo cuando sea posible y cómo controlarlo y estimar sus efectos cuando el sesgo es inevitable.

Por otro lado, los errores aleatorios, consecuencia del factor azar, son inherentes a todas las observaciones. Se pueden minimizar, pero nunca evitar completamente. Esta fuente de error se denomina «aleatoria», porque, por término medio, es tan probable que dé como resultado valores observados situados a un lado del valor real, como valores del otro lado.

Muchos médicos tienden a subestimar la importancia del azar en lo que concierne a los sesgos cuando interpretan los datos, quizás porque las estadísticas son cuantitativas y parecen ser tan definitivas. Básicamente, uno puede pensar, «si las conclusiones estadísticas son sólidas, algún pequeño sesgo no puede causar demasiados problemas». Sin embargo, cuando los datos presentan sesgos, es imposible que la elegancia estadística pueda salvarlos. Como señaló un estudiante, quizá adoptando una posición extrema, «un estudio bien diseñado y realizado minuciosamente suele proporcionar resultados que son obvios sin un análisis formal, y si hay defectos considerables en el diseño o ejecución, un análisis formal no será de ninguna ayuda» (2).

En este capítulo se trata el azar en el contexto de un ensayo clínico controlado porque es una manera sencilla de presentar los conceptos. Sin embargo, las estadísticas son un elemento de toda investigación clínica, cuando se hacen inferencias sobre poblaciones en función de la información obtenida de las muestras. En consecuencia, siempre existe una posibilidad de que la muestra específica de los pacientes en estudio, aunque se seleccionen de un modo sin sesgos, no se parezca a la de la población de pacientes en su totalidad. La estadística ayuda a estimar cómo se aproximan a la realidad las observaciones de las muestras.

booksmedicos.org

DOS ENFOQUES DEL AZAR

Para evaluar el papel que desempeña el azar en las observaciones clínicas, se utilizan dos enfoques generales.

El primero, denominado **verificación de la hipótesis,** se encarga de comprobar si un efecto (diferencia) está presente o no, mediante el uso de pruebas estadísticas para examinar la hipótesis (la «hipótesis nula»), y de que no hay diferencia. Este es el método tradicional para evaluar el papel del azar y se relaciona con los famosos «valores de *P*». Se ha hecho popular desde que se introdujo la metodología estadística a principios del siglo XX. El enfoque de verificación de la hipótesis da lugar a dos conclusiones dicotómicas: el efecto está presente o no hay pruebas suficientes para concluir si un efecto está presente.

El otro enfoque, denominado **estimación,** emplea métodos estadísticos para calcular el rango de valores que es probable que incluya el valor real de una tasa, medida del efecto o rendimiento de una prueba. Últimamente, este enfoque se ha hecho más popular y muchas revistas médicas lo prefieren, al menos para informar de los efectos principales, por las razones que se describen a continuación.

VERIFICACIÓN DE LA HIPÓTESIS

Normalmente, las conclusiones principales de un ensayo se expresan en términos dicotómicos, como un tratamiento nuevo que es mejor o peor que el tratamiento habitual, que corresponde a resultados que son **«estadísticamente significativos»** (improbable que sean por pura casualidad) o no lo son. Las conclusiones estadísticas pueden relacionarse con la realidad de cuatro maneras (fig. 11-1).

Dos de las cuatro posibilidades dan lugar a conclusiones correctas: *i)* el tratamiento nuevo es realmente mejor y esta es la conclusión del estudio, y *ii)* los tratamientos tienen realmente efectos similares y el estudio concluye que es improbable que haya una diferencia.

booksmedicos.org

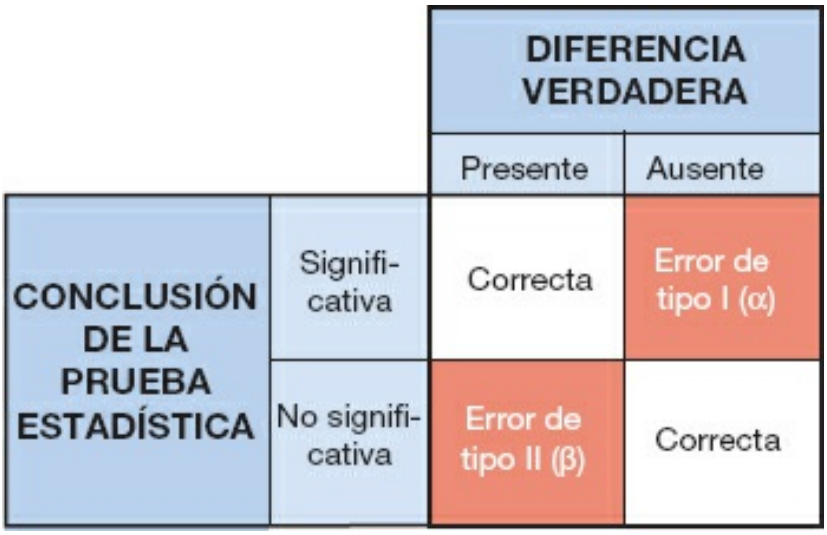


Figura 11-1 ■ Relación entre los resultados de una prueba estadística y la diferencia verdadera entre dos grupos de tratamiento. (*Ausente* es una simplificación; realmente significa que la diferencia verdadera no es mayor que una cantidad especificada.)

Resultados falsos negativos y falsos positivos

También existen dos maneras de equivocarse. El tratamiento nuevo y la atención habitual, de hecho, pueden tener efectos similares, pero se concluye que el tratamiento nuevo es más eficaz. Los errores de este tipo, que comportan una conclusión «positiva-falsa» de que el tratamiento es eficaz, se denominan **error de tipo I** o **error α,** probabilidad de manifestar que existe una diferencia en los efectos de los tratamientos cuando no la hay.

Por otro lado, el tratamiento nuevo puede ser más eficaz, pero el estudio concluye que no lo es. Esta conclusión «negativa-falsa» se denomina **error de tipo II** o **error β,** probabilidad de expresar que no hay ninguna diferencia en los efectos de los tratamientos cuando sí que la hay. «Ninguna diferencia» es una manera simplificada de afirmar que la verdadera diferencia es improbable que sea superior a un cierto tamaño, que se considera demasiado pequeño como para tener una consecuencia útil. No es posible establecer que no hay diferencia alguna entre los dos tratamientos.

La figura 11-1 es similar a la tabla de 2 × 2 que compara los resultados de una prueba diagnóstica con el diagnóstico real (v. cap. 8). Aquí, la «prueba» es la conclusión de un ensayo clínico basado en la comprobación estadística de los resultados de una muestra de pacientes. El «patrón de referencia» es la verdadera diferencia de los tratamientos que se comparan:

booksmedicos.org

si se pudiera establecer, por ejemplo, observando a todos los pacientes que sufren la enfermedad o un gran número de muestras de estos pacientes. El error de tipo I es análogo a un resultado de prueba falso positivo y el error de tipo II a uno falso negativo. En ausencia de sesgos, la variación aleatoria es responsable de las dudas de una conclusión estadística.

Dado que la variación aleatoria forma parte de todas las observaciones, es más sencillo preguntarse si el azar es responsable de los resultados. Mejor dicho, se trata de saber qué probabilidad hay de que la variación aleatoria justifique los resultados en las condiciones específicas del estudio. La probabilidad de error debido a la variación aleatoria se calcula por medio de la **estadística deductiva o inferencial,** una ciencia cuantitativa que, dadas determinadas suposiciones sobre las propiedades matemáticas de los datos, es la base para calcular la probabilidad de que los resultados se hubiesen obtenido sólo por azar.

La estadística es un campo especializado con su propia jerga (p. ej., hipótesis nula, varianza, regresión, potencia y modelos) que es desconocida por muchos médicos. Sin embargo, dejando a un lado la complejidad genuina de los métodos estadísticos, una persona desconocedora de la materia debe considerar la estadística deductiva como un medio útil para un fin. La **verificación estadística** es un medio por el que se calculan los efectos de la variación aleatoria.

Las dos secciones siguientes tratan los errores de tipo I y de tipo II, y pretenden situar la verificación de la hipótesis, como se utiliza para calcular las probabilidades de estos errores, en el contexto.

Concluyendo sobre el buen funcionamiento de un tratamiento

La mayor parte de la estadística que se encuentra en la bibliografía médica concierne la posibilidad de un error de tipo I, y se expresa por el conocido **valor de *P*,** que es un valor cuantitativo de la probabilidad de que las diferencias de los efectos del tratamiento en el estudio específico hubieran podido producirse sólo por azar, suponiendo que, de hecho, no haya diferencia alguna entre los grupos. Otra forma de expresarlo es que *P* es una respuesta a la pregunta, «Si no hubiera ninguna diferencia entre los tratamientos y el ensayo se repitiera muchas veces, ¿qué proporción de los ensayos concluiría que la diferencia entre los dos tratamientos es, como mínimo, tan grande como la detectada en el estudio?»

booksmedicos.org

En esta presentación, el valor de *P* se denomina P_α para distinguirlo de los cálculos del otro tipo de error derivado de la variación aleatoria, errores de tipo II, a los que nos referiremos como P_β. Cuando en la bibliografía científica encontramos una simple *P*, se refiere a P_β.

El tipo de error que calcula P_α se aplica cuando se concluye que un tratamiento es más eficaz que otro. Si se concluye que P_α sobrepasa algún límite (v. más adelante), y no existe ninguna diferencia estadística entre los tratamientos, el valor específico de P_α no es tan relevante; en esta situación, se aplica el valor de P_β (probabilidad de error de tipo II).

Valores de *P* dicotómicos y exactos

Se ha convertido en una costumbre el hecho de atribuir una importancia especial a los valores de *P* inferiores a 0,05 porque existe un consenso en que una posibilidad menor que 1 de 20 es un riesgo de error lo suficientemente pequeño. De hecho, una posibilidad de 1 en 20 es tan pequeña que es razonable concluir que esta incidencia es improbable que se produzca sólo por casualidad. Se podría obtener al azar, y existe una posibilidad entre 20 de que así sea, pero es improbable.

Las diferencias asociadas a P_α inferiores a 0,05 se denominan **significación estadística.** Sin embargo, establecer un límite en 0,05 es totalmente arbitrario. Algunas personas pueden aceptar valores superiores o insistir en inferiores, dependiendo de las consecuencias de una conclusión positiva falsa en una situación determinada. Por ejemplo, uno puede estar dispuesto a aceptar una posibilidad mayor de obtener un resultado positivo falso de una prueba estadística si la enfermedad es grave, no hay ningún tratamiento eficaz en ese momento ni tratamiento nuevo seguro. Por otro lado, uno puede ser reticente a aceptar un resultado falso positivo de una prueba si la atención habitual es eficaz y el tratamiento nuevo es peligroso o mucho más costoso. Este razonamiento es similar al que se aplica a la importancia del resultado falso positivo y falso negativo de una prueba diagnóstica (v. cap. 8).

Para satisfacer varias opiniones sobre lo que es probable y lo que no, algunos investigadores notifican las probabilidades exactas de *P* (p. ej., 0,03; 0,07; 0,11; etc.), en lugar de agruparlas sólo en dos categorías (≤ 0,05 o > 0,05). En consecuencia, los usuarios son libres de aplicar sus

booksmedicos.org

preferencias sobre lo que es estadísticamente significativo. Sin embargo, los valores de *P* mayores de 1 en 5 se suelen indicar sencillamente $P > 0{,}20$, porque casi todo el mundo está de acuerdo en que una probabilidad de error de tipo I superior a 1 en 5 es tan alta que resulta inadmisible. Del mismo modo, con valores de *P* extremadamente bajos (como $P < 0{,}001$), es muy poco probable que el azar pueda detectar una diferencia, y mediante la descripción más precisa de esta posibilidad se transmite algo más de información.

Otro enfoque es aceptar la supremacía de $P \leq 0{,}05$ y describir los resultados que se acercan a este estándar, en términos como «casi estadísticamente significativa», «no alcanza la significación estadística», «marginalmente significativa» o una «tendencia». Estos términos cargados de valor sugieren que el hallazgo debería haber sido estadísticamente significativo, pero que por alguna molesta razón no lo fue. Es mejor simplemente mencionar el resultado y el valor exacto de *P* (o la estimación puntual y el intervalo de confianza [IC], v. más adelante) y dejar que el lector decida por sí mismo si el azar podría haber afectado al resultado.

Significación estadística e importancia clínica

Una diferencia estadísticamente significativa, independientemente de lo pequeño que sea el valor de *P*, no significa que la diferencia sea clínicamente importante. Un valor de $P < 0{,}0001$, si resulta de un estudio bien diseñado, transmite con un grado de confianza elevado que la diferencia realmente existe, pero no explica nada sobre la magnitud de la diferencia ni sobre la importancia clínica de esta. De hecho, las diferencias insignificantes pueden ser estadísticamente muy significativas si se han estudiado bastantes pacientes.

Ejemplo

Se desarrolló el fármaco donepezilo, un inhibidor de la colinesterasa, se desarrolló para el tratamiento de la enfermedad de Alzheimer. En un ensayo controlado y aleatorizado para determinar si el fármaco provocaba mejorías, se asignó a 565 pacientes con esta enfermedad, de forma aleatoria, donepezilo o un placebo (3). La significación estadística de algunos criterios de valoración fue impresionante: tanto el examen mini-mental cognoscitivo (MEC) como la escala Bristol de las actividades de la vida diaria (BADLS) fueron estadísticamente diferentes en $P < 0{,}0001$. Sin embargo, las diferencias reales eran pequeñas: 0,8 en una escala

booksmedicos.org

de 30 puntos para la MEC, y 1 en una escala de 60 puntos para la BADLS. Además, otros resultados, que representaban mucho más el peso de la enfermedad y la atención de estos pacientes, fueron similares en los grupos de donepezilo y placebo. Estos comportaban la atención institucional y la progresión de la enfermedad (ambos variables principales de valoración), así como síntomas conductuales y psicológicos, atención de profesionales sanitarios, costos formales de la asistencia, tiempo de los profesionales sin remuneración, y los efectos adversos o la muerte. Los autores concluyeron que los beneficios del donepezilo estaban «muy por debajo del umbral mínimo relevante».

Por otro lado, los valores de *P* muy insignificantes pueden ser consecuencia de estudios que muestran efectos del tratamiento sólidos si se estudian pocos pacientes.

Pruebas estadísticas

Las **pruebas estadísticas** se utilizan para calcular la probabilidad de un error de tipo I. La prueba que se aplica a los datos para obtener un resumen numérico se denomina prueba estadística. Posteriormente, esta cifra se compara con la distribución del muestreo para obtener la probabilidad de un error tipo I (fig. 11-2). La distribución es bajo la **hipótesis nula,** la proposición de que no hay ninguna diferencia en los resultados entre los grupos de tratamiento. Este concepto es por razones matemáticas, y no porque «ninguna diferencia» sea la hipótesis científica de los investigadores que dirigen el estudio. Uno termina rechazando la hipótesis nula (llegando a la conclusión de que hay una diferencia) o no logrando rechazarla (concluyendo que no hay pruebas suficientes a favor de la diferencia). Obsérvese que el hecho de no encontrar significación estadística no quiere decir que haya diferencias. La verificación estadística es incapaz de establecer que no haya diferencia alguna.

En la tabla 11-1 aparecen algunas pruebas estadísticas que se utilizan con mucha frecuencia. La validez de muchas pruebas depende de ciertas conjeturas sobre los datos; una hipótesis típica es que estos tienen una distribución normal. Si los datos no cumplen estas conjeturas, el valor de *P* resultante puede ser erróneo. Otras pruebas estadísticas, denominadas **pruebas no paramétricas,** no consideran ninguna hipótesis sobre la distribución subyacente de los datos. Cualquier manual de texto de bioestadística explica cómo se obtienen y se calculan estas pruebas estadísticas, y las conjeturas en las que se basan.

booksmedicos.org

Tabla 11-1

Algunas pruebas estadísticas utilizadas habitualmente en la investigación clínica

Prueba	Cuándo se utiliza
Para comprobar la significación estadística de una diferencia	
Ji al cuadrado (χ^2)	Entre dos o más proporciones (cuando hay un gran número de observaciones)
Prueba exacta de Fisher	Entre dos proporciones (cuando hay un número pequeño de observaciones)
U de Mann-Witney	Entre dos medianas
t de Student	Entre dos medias
Prueba de la *F*	Entre dos o más medias
Para describir el grado de relación	
Coeficiente de regresión	Entre una variable independiente (predictiva) y una dependiente (de respuesta)
r de Pearson	Entre dos variables
Para modelar los efectos múltiples variables	
Regresión logística	En un resultado dicotómico
Riesgos proporcionales de Cox	En tiempo hasta el evento dicotómico

La prueba de ji cuadrado (χ^2) para datos nominales se entiende más fácilmente que la mayoría y se puede utilizar para ejemplificar cómo funciona una prueba estadística. El grado en que los valores observados se alejan de lo esperado, si no hubiera un efecto del tratamiento, se utiliza para calcular el valor de *P*.

booksmedicos.org

Figura 11-2 ■ Verificación estadística.

Ejemplo

El paro cardíaco extrahospitalario tiene un pronóstico desalentador. Los estudios en animales sugieren que la hipotermia podría mejorar los resultados neurológicos. Para probar esta hipótesis en humanos, 77 pacientes que permanecieron inconscientes después de la reanimación por un paro cardíaco extrahospitalario fueron aleatorizados al enfriamiento (hipotermia) o al tratamiento habitual (4). El resultado principal fue la supervivencia a la salida hospitalaria con una función neurológica relativamente buena.

Tasas observadas

	Superviviente con función neurológica		
	Sí	No	Total
Hipotermia	21	22	43
Asistencia estándar	9	25	34
Total	30	47	77

Las tasas de éxito fueron del 49 % en pacientes tratados con hipotermia y del 26 % en los tratados con el manejo habitual. ¿Qué tan probable sería para un estudio de este tamaño observar una diferencia en las tasas tan grande como esta o mayor si realmente no existiera una diferencia en la efectividad? Esto depende hasta qué punto los resultados observados se alejan de lo que se habría esperado si los tratamientos tuvieran una efectividad similar y sólo la variación aleatoria hubiera explicado las tasas diferentes. Si el tratamiento no tuviera ningún efecto sobre el resultado, aplicando la tasa adecuada para los pacientes en su totalidad (30/77 = 39 %) al número de pacientes en cada grupo de tratamiento, se obtiene el número esperado de sucesos en cada grupo.

booksmedicos.org

Tasas esperadas

	Sucesos		
	Sí	No	Total
Hipotermia	16,75	26,25	43
Asistencia estándar	13,25	20,75	34
Total	30	47	77

La estadística χ^2 es el cuadrado de las diferencias entre lo observado y lo esperado, dividido entre lo esperado, resumida en las cuatro casillas:

$$\chi^2 = \sum \frac{(\text{Número observado} - \text{Número esperado})^2}{\text{Número esperado}}$$

La diferencia entre los números observados y los esperados, si no hubiera efecto del tratamiento, determina la magnitud del estadístico χ^2. Al ser al cuadrado, no importa si los índices observados sobrepasan los esperados o caen cerca de ellos. Al dividir la diferencia al cuadrado de cada casilla por el número esperado, la diferencia para esa casilla se ajusta al número de pacientes de la misma.

El estadistico χ^2 para los siguientes datos es:

$$\frac{(21-16{,}75)^2}{16{,}75} + \frac{(9-13{,}25)^2}{13{,}25} + \frac{(22-26{,}25)^2}{26{,}75} + \frac{(25-20{,}75)^2}{20{,}75} = 4{,}0$$

A continuación se compara el valor de χ^2 (disponible en libros y programas informáticos) en una tabla en la que se relacionan los valores con las probabilidades según el número de casillas, con el fin de obtener la probabilidad de que la χ^2 sea de 4,0. Es evidente que, cuanto mayor sea χ^2, menos probabilidades hay de que justifique las diferencias que se observan. El valor de *P* resultante para una prueba estadística de ji cuadrado de 4,0 y una tabla de 2 × 2 fue de 0,046, que es la probabilidad de una conclusión falsa positiva sobre la diferencia en el efecto del tratamiento. Es decir, los resultados cumplen con el criterio convencional de significación estadística, $P \leq 0{,}05$.

Cuando se utilizan pruebas estadísticas, el enfoque habitual es poner a prueba la probabilidad de que una intervención es más o menos eficaz que la otra en un grado estadísticamente importante. En esta situación, la prueba se denomina de **dos colas,** en referencia a las dos colas de una curva con forma de campana que describe la variación aleatoria en las

booksmedicos.org

diferencias entre los grupos de tratamiento de igual valor, donde las dos colas de la curva incluyen resultados estadísticamente improbables que favorecen uno u otro tratamiento. Algunas veces existen razones de peso para creer que un tratamiento sólo podría ser mejor o peor que el otro, en cuyo caso se utiliza una prueba de **una cola,** donde todo el error de tipo I (5 %) está en una de las colas, lo que facilita alcanzar la significación estadística.

Conclusión sobre el mal funcionamiento de un tratamiento

Algunos ensayos son incapaces de inferir que un tratamiento es mejor que otro. El riesgo de obtener un resultado falso negativo es especialmente importante en los estudios con relativamente pocos pacientes o episodios de resultado. Entonces surge la pregunta: ¿qué posibilidad cabe de que el resultado sea falso negativo (error de tipo II o error β)? ¿Es posible que los descubrimientos de dichos ensayos hayan distorsionado la realidad porque los resultados de estos estudios concretos han tenido la mala suerte de aparecer de una manera poco probable?

Ejemplo

Uno de los ejemplos del capítulo 9 trató de un ensayo aleatorizado controlado de los efectos cardiovasculares de la adición de niacina frente a placebo en pacientes con anormalidades en los lípidos que ya estaban tomando estatinas (5). Fue un ensayo «negativo»: los resultados principales se presentaron en el 16,4 % de los pacientes que tomaron niacina y en el 16,2 % de los que tomaron placebo, y los autores concluyeron que «no hubo un mayor beneficio clínico al añadir niacina al tratamiento con estatinas». La pregunta estadística asociada con esta afirmación es: ¿Qué tan probable es que el estudio no encontrara un beneficio cuando realmente sí existe? Después de todo, sólo había unos cuantos cientos de acontecimientos de resultado cardiovascular en el estudio, por lo que la participación del azar podría haber ocultado los efectos del tratamiento. La figura 11-3 muestra las curvas del tiempo hasta el episodio para el resultado principal en los dos grupos de tratamiento. Los pacientes en los grupos de niacina y de control tuvieron curvas muy similares durante todo el seguimiento, por lo que el efecto protector de la niacina es inverosímil.

La presentación visual de los resultados negativos puede ser convincente. De forma alternativa, se pueden examinar los intervalos de confianza (v. más adelante) y determinar en buena medida si el estudio fue lo suficientemente grande como para descartar las diferencias clínicamente

booksmedicos.org

importantes si hubieran existido.

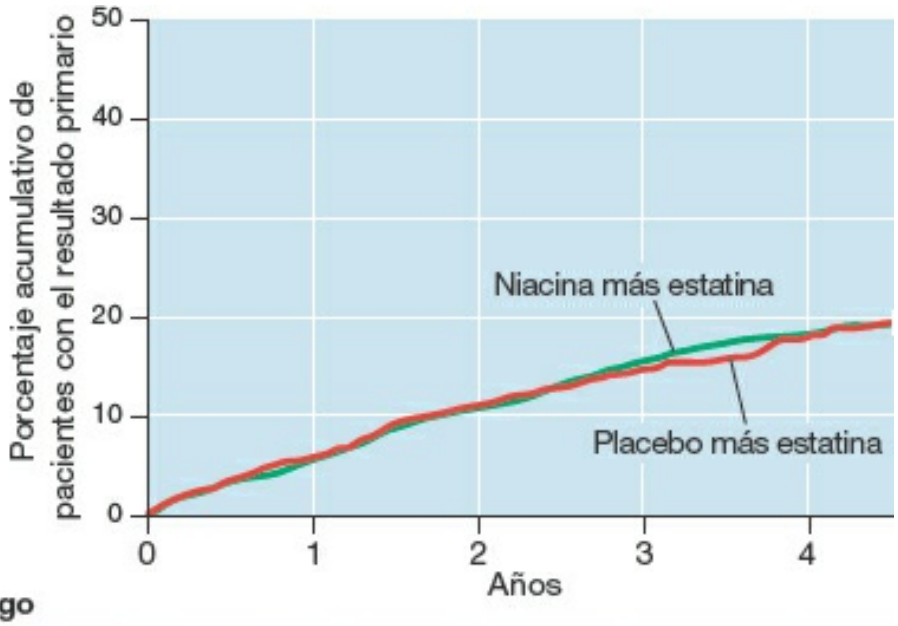
Por supuesto, también se deben considerar otras causas de resultados falsos negativos además del azar: causas biológicas como un seguimiento demasiado breve o una dosis demasiado baja de niacina, además de las limitaciones del estudio como el incumplimiento o los episodios de resultado omitidos.

Los errores de tipo II han acaparado menos la atención que los errores de tipo I por varias razones. Son más difíciles de calcular. Asimismo, la mayoría de los profesionales prefieren sencillamente las cosas que funcionan y los resultados negativos no son bien recibidos. Es menos probable que los autores envíen a las revistas estudios con resultados negativos y, si estos se notifican, los autores pueden preferir hacer hincapié en los subgrupos de pacientes en los que se encuentran diferencias del tratamiento. Además, los autores pueden subrayar razones, aparte del azar, para explicar por qué las diferencias verdaderas pueden haberse omitido. Cuál es el motivo por el que no se ha considerado la probabilidad de un error de tipo II, es la pregunta fundamental que se debe formular cuando los resultados de un estudio no indican diferencias.

¿CUÁNTOS PACIENTES SE NECESITAN PARA UN ESTUDIO?

Imagínese que está leyendo un ensayo clínico en el que se compara un tratamiento nuevo con el habitual y no encuentra diferencia alguna. Usted es consciente de que la variación aleatoria puede ser la fuente de cualquier diferencia que se observe, y se pregunta si la cantidad de pacientes (tamaño muestral) de este estudio es lo suficientemente grande como para hacer que el azar sea una explicación poco probable de los resultados que se han obtenido. De forma alternativa, usted puede estar planeando realizar un estudio de este tipo y tener la misma pregunta. De cualquier manera, es necesario comprender cuántos pacientes se necesitarán para comparar de forma adecuada los efectos de los dos tratamientos.

booksmedicos.org



Número en riesgo					
Niacina más estatina	1 718	1 606	1 366	903	428
Placebo más estatina	1 696	1 581	1 381	910	436

Figura 11-3 ■ Ejemplo de un ensayo «negativo». (Redibujado con autorización The AIM-HIGH Investigators. Niacin in patients with low HDL cholesterol levels receiving intensive statin therapy. N Engl J Med 2011;365:2255-2267.)

Potencia estadística

Se denomina **potencia estadística** o **poder estadístico** del ensayo a la probabilidad de que un ensayo detecte una diferencia estadísticamente significativa cuando la diferencia realmente existe. La potencia y P_β son maneras complementarias de expresar el mismo concepto

$$\text{Potencia estadística} = 1 - P_\beta$$

La potencia es análoga a la sensibilidad de una prueba diagnóstica. Se dice que un estudio es potente si hay muchas probabilidades de que se detecten diferencias cuando los tratamientos realmente tienen efectos diferentes.

Requisitos para calcular el tamaño de la muestra

Desde el punto de vista de la verificación de la hipótesis de los datos nominales (recuento), el **tamaño muestral** apropiado depende de cuatro características del estudio: la magnitud de la diferencia en los resultados

booksmedicos.org

entre los grupos de tratamiento; P_{α} y P_{β} (la probabilidad de obtener un resultado falso positivo y falso negativo que se está dispuesto a aceptar) y la tasa del resultado del estudio.

Estos determinantes de un tamaño de muestra apropiado deben tenerse en cuenta cuando los investigadores planifican un estudio, para garantizar que este tendrá la suficiente potencia estadística para producir resultados significativos. En la medida en que los investigadores no lo hayan ejecutado correctamente, o algunas de sus hipótesis fueron inexactas, los lectores deben considerar los mismos problemas cuando interpreten los resultados del estudio.

Magnitud del efecto

El tamaño muestral depende de la magnitud de la diferencia que se debe detectar. Uno es libre de buscar diferencias de cualquier magnitud y, por supuesto, se espera poder detectar incluso diferencias muy pequeñas. No obstante, en muchos pacientes se precisa detectar pequeñas diferencias, siendo todo lo demás igual. Así pues, es mejor sólo preguntarse cuál es el número de pacientes necesario para que se pueda detectar la mínima mejoría que sea clínicamente importante. Por otro lado, si uno se interesa sólo en detectar grandes diferencias entre los grupos tratados y los de control (es decir, efectos del tratamiento sólidos) se han de estudiar pocos pacientes.

Error de tipo I

El tamaño muestral también está relacionado con el riesgo de un error de tipo I (al concluir que el tratamiento es eficaz cuando no lo es). La probabilidad aceptable para un riesgo de este tipo es un juicio valorativo. Si se está preparado para aceptar las consecuencias de que haya una elevada posibilidad de error, concluyendo que el tratamiento es eficaz, se puede llegar a conclusiones con relativamente pocos pacientes. Por otro lado, si se desea afrontar sólo un riesgo mínimo de equivocarse en este sentido, se necesita una mayor cantidad de pacientes. Como ya se comentó anteriormente, P_{α} se establece en 0,05 (1 en 20) o, a veces, en 0,01 (1 en 100).

Error de tipo II

El riesgo escogido de tipo II es otro determinante del tamaño muestral.

booksmedicos.org

Una probabilidad aceptable de este error también es un juicio que se puede emitir y cambiar libremente según las preferencias personales. La probabilidad de P_β se suele fijar en 0,20; un 20 % de posibilidades de omitir diferencias verdaderas en un estudio concreto. Los errores de tipo II tradicionales son mucho mayores que los errores de tipo I, y reflejan un valor mayor que normalmente se sitúa en asegurar que un efecto está realmente presente, cuando se dice que está.

Características de los datos

La potencia estadística de un estudio también está determinada por la naturaleza de los datos. Cuando el resultado se expresa con recuento o proporciones de episodios o del período de tiempo para el episodio, la potencia estadística depende del índice de episodios: cuanto mayor sea la cantidad de episodios, mayor será la potencia estadística para un número determinado de personas en situación de riesgo. Como Peto y cols. Expusieron (6):

> En ensayos clínicos de pronóstico para que se produzca la muerte (o el tiempo para cualquier otro «episodio» concreto como recaída, metástasis, primera trombosis, enfermedad cerebrovascular, recidiva o pronóstico para la muerte por una causa específica), la capacidad del ensayo para distinguir entre los beneficios de dos tratamientos depende del número de pacientes *fallecidos* (o que sufren un episodio relevante), y no de la cantidad de pacientes *inscritos*. Un estudio de 100 pacientes, 50 de los cuales fallecieron, es casi tan sensible como un estudio de 1 000 pacientes, 50 de los cuales perecieron.

Tabla 11-2

Determinantes del tamaño muestral

	Determinado por			
		Tipos de datos		
	Investigador	Medias		Recuentos
El tamaño de la muestra varía de acuerdo a:	$\frac{1}{\text{Tamaño del efecto}, P_\alpha, P_\beta}$	$\frac{1}{\text{Tasa de resultados}}$	O	Variabilidad

Si los datos son continuos, como la presión arterial o el colesterol sérico, el grado en que los pacientes varían entre sí afecta a la potencia. Cuanto mayor sea la variación de un paciente a otro respecto a la característica que se mide, más difícil es estar seguro de que las diferencias observadas (o la ausencia de diferencia) entre los grupos no se deben a esta

booksmedicos.org

variación, sino más bien a un efecto del tratamiento verdadero.

Al diseñar un estudio, los investigadores seleccionan el menor efecto del tratamiento que sea clínicamente significativo (los efectos de un tratamiento más grandes serán más fáciles de detectar), y los errores de tipo I y de tipo II que están dispuestos a aceptar. También obtienen estimaciones de los datos de los resultados de índices de episodios o la variación entre pacientes. Es posible diseñar estudios que maximicen la potencia para un tamaño muestral determinado, por ejemplo escogiendo pacientes con una elevada tasa de episodios o características similares, con tal que se adapten al tema de la investigación.

Interrelaciones

En la tabla 11-2 se resumen las relaciones entre las cuatro variables que conjuntamente determinan la potencia estadística. Las variables se pueden intercambiar una con otra. En general, para un determinado número de pacientes en el estudio, existe una interrelación entre los errores de tipo I y los de tipo II. Si todo lo demás es igual, cuanto más se esté dispuesto a aceptar un tipo de error, menos necesario será arriesgar el otro. Ningún tipo de error es intrínsecamente peor que el otro. Evidentemente, es posible reducir ambos errores si se aumenta la cantidad de pacientes, los episodios del resultado son más frecuentes, disminuye la variabilidad o se busca un efecto del tratamiento mayor.

Con los siguientes ejemplos, se ilustra la relación entre la magnitud del efecto del tratamiento y la cantidad de pacientes que se necesitan para un ensayo con niveles tradicionales de P_{α} y P_{β}. Uno representa una situación en la que fue suficiente una cantidad de pacientes relativamente pequeña, y el otro es un ejemplo en el que una gran cantidad de pacientes fue demasiado pequeña.

Ejemplo

Un tamaño muestral pequeño fue adecuado

Durante muchos siglos, el escorbuto, un síndrome causado por deficiencia de vitamina C caracterizado por gingivorragia y hemartrosis, produciendo discapacidad y muerte, fue un grave problema para los marineros. Mientras se estaba a bordo de un barco en 1747, James

booksmedicos.org

Lind estudió el tratamiento del escorbuto en uno de los primeros ensayos controlados sobre tratamiento (7). Él confinó a 12 marineros gravemente afectados en la enfermería del barco y los dividió en 6 grupos de tratamiento con 2 marineros cada uno. Los tratamientos fueron naranjas y limones, sidra, elixir de vitriolo, vinagre, agua de mar o nuez moscada. Con excepción del tratamiento, los marineros permanecieron en el mismo sitio y comieron la misma comida. Seis días después de comenzar el tratamiento, los dos marineros que consumieron naranjas y limones, que contienen grandes cantidades de vitamina C, habían mejorado mucho: uno estaba listo para cumplir con sus labores y el otro estaba lo suficientemente sano como para atender al resto. El resto, a quienes se les administró otros tratamientos, no mejoraron. El valor de P (que no fue calculado por Lind) para 2 de 2 éxitos contra 10 de 10 fracasos fue 0,02 (considerando que Lind hubiera creído desde el principio que las naranjas y los limones eran la intervención de interés y que las otras cinco intervenciones eran controles inactivos).

Un tamaño muestral grande fue inadecuado

Los niveles séricos bajos de vitamina D pueden ser un factor de riesgo para el cáncer colorrectal, pero los resultados de los estudios sobre este tema han mostrado inconsistencias. En 1985, los investigadores formaron (originalmente, para otro propósito) una cohorte de 29 133 hombres finlandeses de 50 a 69 años y midieron sus niveles basales de vitamina D en suero (8). A partir del National Cancer Registry se identificaron los carcinomas de colon y recto incidentales. Después de 12 años de seguimiento, se desarrollaron 239 carcinomas de colon y 192 carcinomas de recto en la cohorte. Después de ajustar los factores de confusión, los niveles séricos de vitamina D se asociaron positivamente con la incidencia de carcinoma de colon y se asociaron negativamente con la incidencia de carcinoma rectal, aunque ninguno de estos hallazgos fue estadísticamente significativo.

Para la mayor parte de las cuestiones terapéuticas con las que nos encontramos hoy en día, hace falta un tamaño muestral extremadamente grande. El valor de tratamientos potentes y espectaculares, como los antibióticos para la neumonía o el reemplazo tiroideo para el hipotiroidismo, se estableció mediante la experiencia clínica o estudiando un pequeño número de pacientes. No obstante, estos tratamientos raras veces progresan, y muchos de ellos ya hace tiempo que están establecidos. Hay enfermedades, muchas de las cuales son crónicas y tienen múltiples causas que se relacionan entre sí, para las cuales los efectos del nuevo tratamiento suelen ser pequeños. Por este motivo, es especialmente importante planificar estudios clínicos que sean lo suficientemente grandes para distinguir los efectos reales de los que se producen por el azar.

La figura 11-4 muestra la relación entre el tamaño muestral y las diferencias del tratamiento para tasas basales del efecto de los resultados. En los estudios de menos de 100 pacientes hay pocas posibilidades de detectar diferencias estadísticamente significativas incluso de efectos de tratamientos grandes. Visto de otra manera, es difícil detectar magnitudes de efectos de menos de un 25 %. En la práctica, la potencia estadística se

booksmedicos.org

puede calcular mediante fórmulas, tablas, nomogramas, programas informáticos o sitios web.

ESTIMACIONES PUNTUALES E INTERVALOS DE CONFIANZA

La magnitud del efecto que se observa en un estudio en concreto (p. ej., el efecto de un tratamiento en un ensayo clínico o el riesgo relativo en un estudio de cohortes) se denomina **estimación puntual** del efecto. Es la mejor estimación del estudio, de la verdadera magnitud del efecto, y es el resumen estadístico en el que normalmente se hace más hincapié en los informes de investigación.

No obstante, es poco probable que la magnitud del efecto verdadero sea exactamente la observada en el estudio. Debido a la variación aleatoria, es probable que en cualquier estudio se pueda detectar un resultado superior o inferior al del valor verdadero. Por consiguiente, para la **precisión estadística** de la estimación puntual se necesita un resumen de la medida global, que es la escala de valores que probablemente abarca la magnitud del efecto verdadero.

booksmedicos.org

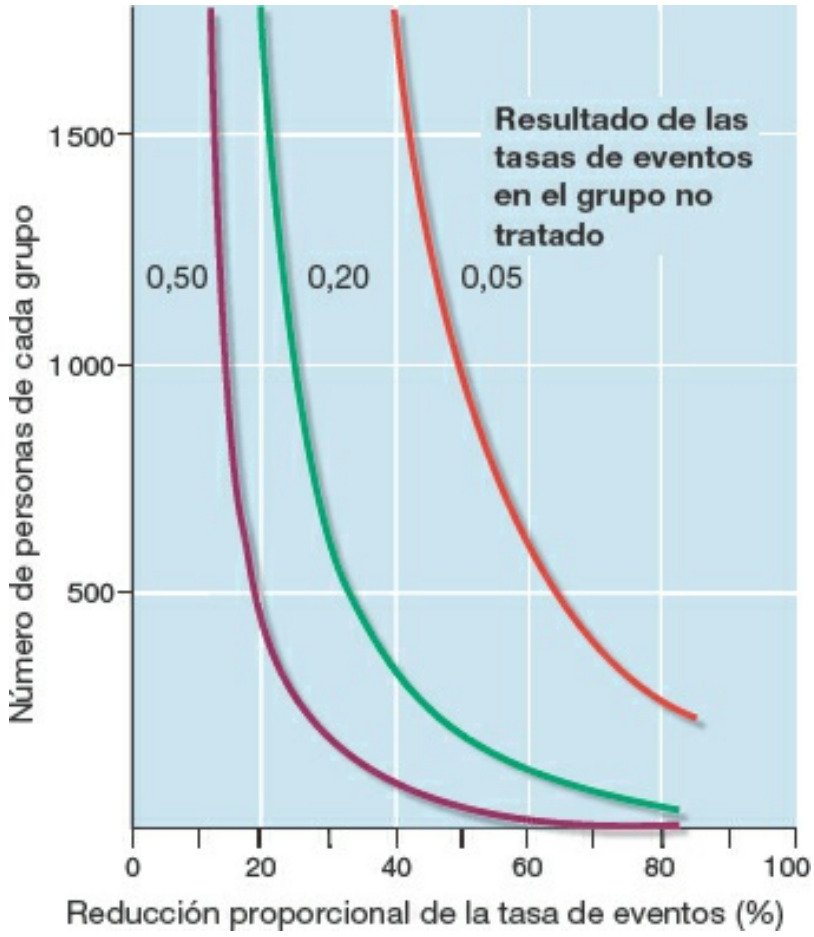


Figura 11-4 ■ Número de personas que se necesita en cada uno de los dos grupos de tratamiento (de igual tamaño), para varios índices de resultados de eventos en el grupo no tratado, para tener un 80 % de posibilidades de detectar una diferencia (*P* = 0,05) en la reducción de los índices de resultados de eventos en pacientes tratados frente a pacientes no tratados. (Calculados a partir de las fórmulas de Weiss NS. Clinical epidemiology. The study of the outcome of illness. New York: Oxford University Press, 1986.)

La precisión estadística se expresa como **intervalo de confianza,** normalmente del 95 %, alrededor de la estimación puntual. Los intervalos de confianza se interpretan del siguiente modo: si el estudio no tiene sesgos, hay un 95 % de probabilidades de que el intervalo incluya la magnitud del efecto verdadero. Cuanto más reducido sea el IC, más fiable será la magnitud del efecto verdadero. Es más probable que el valor real se acerque a la estimación puntual, y es menos probable que se aproxime a los límites externos del intervalo, incluso podría estar fuera de estos límites (5 de cada 100 casos). La precisión estadística aumenta con la potencia estadística del estudio.

booksmedicos.org

Ejemplo

La *Women's Health Initiative* incluyó un ensayo controlado y aleatorizado de los efectos de los estrógenos y la progesterona sobre los resultados de enfermedades crónicas en mujeres menopáusicas sanas (9). En la figura 11-5 se muestran el riesgo relativo y los intervalos de confianza para cuatro de estos resultados: enfermedad cerebrovascular, fractura de cadera, neoplasia mamaria y neoplasia endometrial. Las cuatro ilustran varias posibilidades de la interpretación de los intervalos de confianza. El tratamiento con estrógenos más progesterona comportaba un factor de riesgo para el accidente cerebrovascular; la mejor determinación de este riesgo es la estimación puntual, un riesgo relativo de 1,41, pero los datos concuerdan con el riesgo relativo mínimo de 1,07 y máximo de 1,85. Los estrógenos y la progesterona protegieron frente a las fracturas de cadera, y evitaron un máximo del 65 % y un mínimo del 2 % de las fracturas; es decir, los datos concuerdan con el escaso beneficio, aunque es posible que haya una mejora considerable. Probablemente, el riesgo de sufrir neoplasia mamaria aumente y, sin embargo, los datos no muestran ningún efecto (el extremo inferior del IC incluye un riesgo relativo de 1,0). Por último, el estudio no es demasiado informativo sobre la neoplasia endometrial. Los intervalos de confianza son muy grandes, así que no sólo no había ningún riesgo evidente ni beneficio, sino que el cálculo del riesgo era tan impreciso que seguían siendo posibles un riesgo o beneficio considerables.

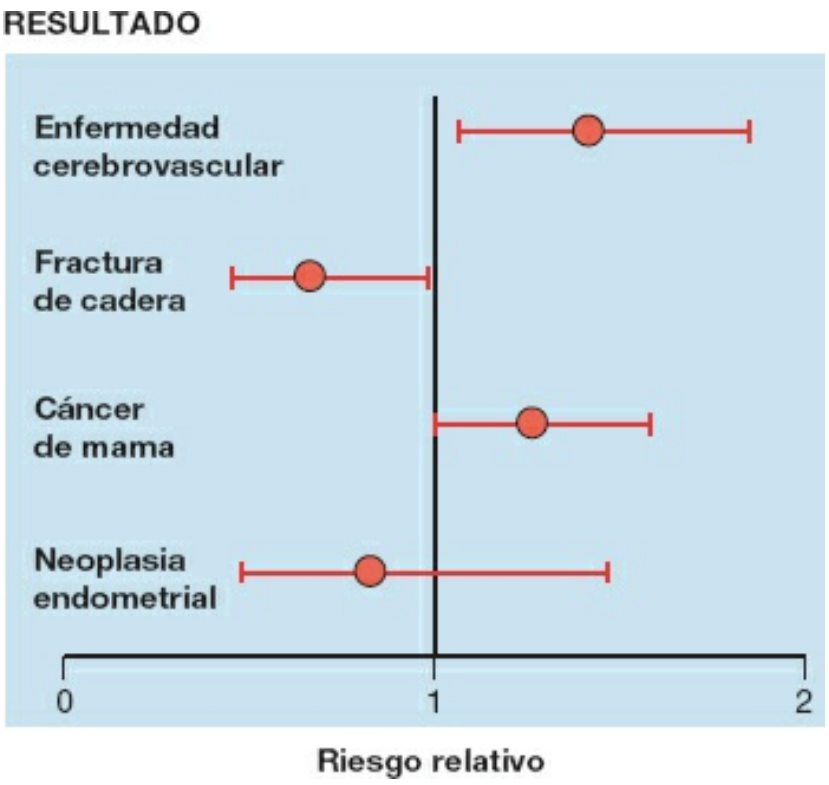


Figura 11-5 ■ Ejemplos de intervalos de confianza. Riesgo relativo e intervalos de confianza en los resultados de la *Women's Health Initiative:* un ensayo controlado y aleatorizado de estrógenos y progesterona en mujeres posmenopáusicas sanas. (De Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal

booksmedicos.org

women. JAMA 2002;288:321-333.)

La significación estadística en el nivel de 0,05 se puede obtener a partir de los intervalos de confianza al 95 %. Si el valor que corresponde a ningún efecto (es decir, un riesgo relativo de 1 o una diferencia de tratamiento de 0) se sitúa fuera de los intervalos de confianza del 95 % para el efecto observado, los resultados son estadísticamente significativos en el nivel 0,05. Si los intervalos de confianza incluyen este punto, los resultados no son estadísticamente significativos.

Los intervalos de confianza presentan ventajas respecto a los valores de *P*. Hacen hincapié donde corresponde, en la magnitud del efecto. Los intervalos de confianza facilitan al lector la visualización del margen de valores plausibles y, de este modo, la decisión de si la magnitud de un efecto que consideran clínicamente significativo concuerda con los datos o es excluido por estos (10). Asimismo, proporcionan información sobre la potencia estadística; si el IC es relativamente ancho, y apenas incluye el valor correspondiente a ningún efecto, los lectores pueden observar que la baja potencia podría haber sido la causa del resultado negativo. Por otro lado, si el IC es estrecho y no incluye un efecto se descarta un gran efecto.

Las estimaciones puntuales y los intervalos de confianza se utilizan para caracterizar la precisión estadística de cualquier tasa (incidencia y prevalencia), rendimiento de la prueba diagnóstica, comparaciones de tasas (riesgo relativo y atribuible) y otros resúmenes estadísticos. Por ejemplo, los estudios han demostrado que un 7,0 % (IC del 95 % de 5,2-9,4) de los adultos tienen antecedentes familiares de cáncer de próstata (11); que la sensibilidad de un análisis de troponina cardiaca de alta sensibilidad (con un punto de corte óptimo) para el síndrome coronario agudo fue del 84,8 % (IC del 95 % de 82,8-86,6) (12), y que el retorno a la actividad habitual después de reparar una hernia inguinal fue más corto en la cirugía laparoscópica que en la cirugía abierta (riesgo relativo de 0,56, IC del 95 % de 0,51-0,61) (13).

Los intervalos de confianza se han convertido en el modo habitual de notificar los resultados principales de las investigaciones clínicas, debido a las muchas ventajas que presentan frente a las pruebas de hipótesis (valor de *P*). Los valores de *P* todavía se utilizan por tradición y comodidad cuando se describen muchos resultados, y no sería práctico incluir intervalos de confianza para todos ellos.

booksmedicos.org

Potencia estadística tras finalizar el estudio

Anteriormente en este capítulo comentamos cómo se realiza el cálculo de la potencia estadística según el abordaje de la verificación de la hipótesis antes de emprender el estudio, para así poder garantizar que habrá los suficientes pacientes para que haya probabilidades de poder detectar un efecto clínicamente significativo, en el caso de que haya alguno. Sin embargo, tras finalizar el estudio, este enfoque resulta menos relevante (14). Ya no se necesita calcular el tamaño muestral, ni los índices del resultado del episodio, ni la variabilidad entre los pacientes porque ya se conocen. En cambio, se debe dirigir la atención a la estimación puntual y a los intervalos de confianza, ya que se puede observar el intervalo de valores que concuerda con los resultados, y si las magnitudes del efecto de interés están dentro del rango o si se excluyen a causa de los datos.

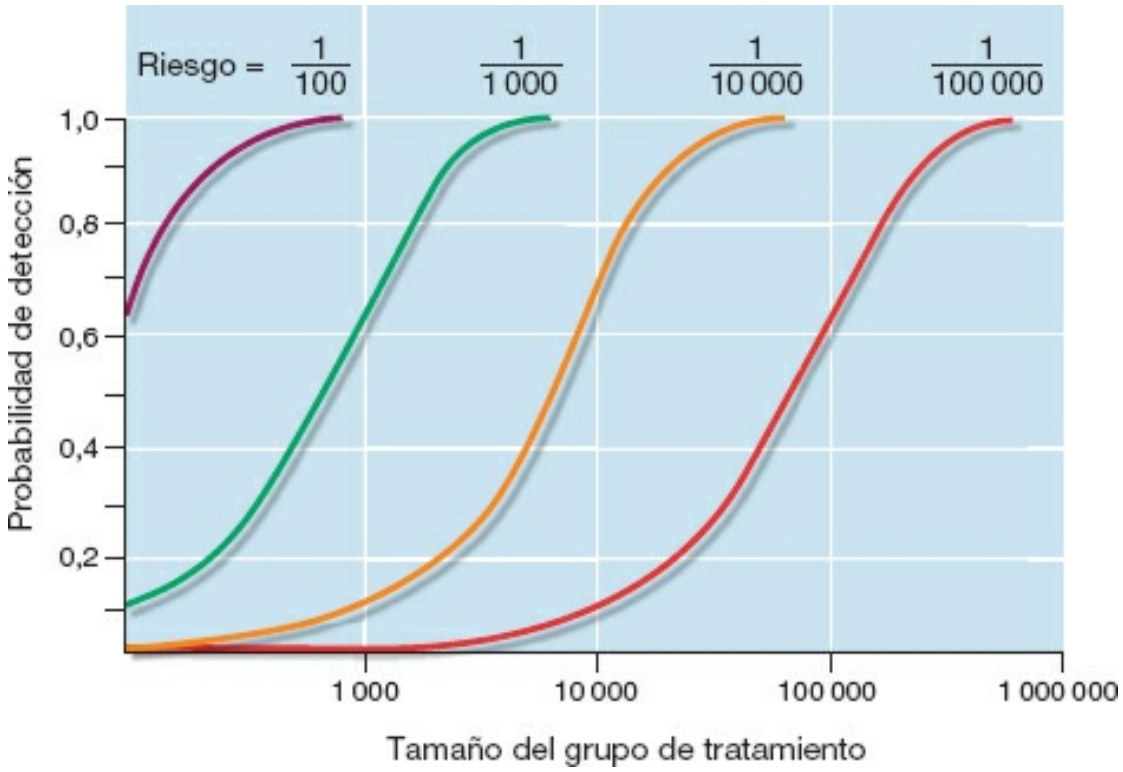


Figura 11-6 ■ Probabilidad de detectar un episodio según la tasa de eventos y el número de personas observadas. (Redibujado con autorización de Guess HA, Rudnick SA. Use of cost effectiveness analysis in planning cancer chemoprophylaxis trials. Control Clin Trials 1983;4:89-100.)

En el estudio de niacina, resumido anteriormente como un ejemplo de un ensayo negativo, el riesgo relativo fue 1,02 y el IC del 95 % fue 0,87-

booksmedicos.org

1,21, lo que quiere decir que los resultados fueron consistentes con un pequeño grado de beneficio o daño. Si esto es significativo depende de la importancia clínica vinculada a una diferencia en las tasas tan grande como la que representa este IC.

DETECCIÓN DE EPISODIOS RAROS

A veces, es importante conocer la probabilidad que hay de que un estudio detecte un episodio relativamente poco frecuente (p. ej., 1/1 000), especialmente si este episodio es grave, como la anemia aplásica o la arritmia potencialmente mortal. Se deben observar muchas personas para tener muchas posibilidades de detectar incluso uno de estos episodios, muchas menos para desarrollar una estimación relativamente estable de la frecuencia. En la mayor parte de las investigaciones clínicas, el tamaño muestral se planifica de modo que sea suficiente para detectar los principales efectos, la respuesta que se busca para la cuestión principal de la investigación. Es probable que el tamaño muestral sea escaso y no llegue a la cantidad necesaria para detectar episodios raros, como por ejemplo efectos secundarios poco frecuentes y complicaciones. Por este motivo, es necesario establecer un enfoque diferente en el que se incluyan muchos más pacientes. Un ejemplo es la farmacovigilancia, que consiste en hacer un seguimiento de miles de personas para observar los efectos secundarios.

En la figura 11-6 se muestra la probabilidad de detectar un episodio en función del número de personas en observación. La regla general es la siguiente: para tener muchas posibilidades de detectar 1/*x* episodios, se deben observar a 3*x* personas (15). Por ejemplo, para detectar por lo menos un episodio si el índice esencial es 1/1 000, se deberían observar 3 000 personas.

COMPARACIONES MÚLTIPLES

Las conclusiones estadísticas de las investigaciones tienen un aura de autoridad que constituye un desafío, especialmente para las personas profanas en la materia. No obstante, como muchos escépticos han sospechado, es posible «engañar con las estadísticas» incluso si la investigación está bien diseñada, las matemáticas son perfectas y las intenciones de los investigadores científicos son irreprochables.

Las conclusiones estadísticas pueden ser erróneas porque el valor de las

booksmedicos.org

pruebas estadísticas depende de la cantidad de preguntas que se consideran en el estudio y cuándo se formularon. Si se realizan muchas comparaciones entre las variables de una gran cantidad de datos, el valor de *P* asociado a cada comparación por separado subestima la frecuencia en que el resultado de esa comparación, entre las otras, es probable que surja por azar. Por improbable que pueda parecer, la interpretación del valor de *P* de una sola prueba estadística depende del contexto en el que se realiza.

Para entender cómo puede suceder esto, considere el siguiente ejemplo.

Ejemplo

Supongamos que se ha realizado un gran estudio en el que hay múltiples subgrupos de pacientes y muchos resultados diferentes. Por ejemplo, puede ser un ensayo clínico del valor de un tratamiento para la arteriopatía coronaria, y los pacientes se dividen en 12 subgrupos clínicamente significativos (p. ej., enfermedad de 1, 2 o 3 vasos; buena o mala función ventricular; presencia o ausencia de arritmias; y varias combinaciones de estas), y se consideran cuatro resultados (p. ej., muerte, infarto de miocardio, insuficiencia cardíaca y angina de pecho). Supongamos también que no existen asociaciones reales entre el tratamiento y el resultado en ninguno de los subgrupos y para ninguno de los resultados. Por último, supongamos que los efectos del tratamiento se evalúan por separado para cada subgrupo y para cada resultado: un proceso que implica muchas comparaciones, en este caso, 12 subgrupos multiplicado por 4 resultados = 48 comparaciones. Con una P_{α} = 0,05 es probable que 1 de cada 20 de estas comparaciones (en este caso, 2 de 3) sea significativa estadísticamente sólo por el azar. En general, si se realizan 20 comparaciones, por término medio, es probable que encontremos una que sea estadísticamente significativa; si se hacen 100 comparaciones, posiblemente unas 5 serán significativas, y así sucesivamente. Por consiguiente, cuando se han hecho muchas comparaciones, algunas pocas excepcionales son suficientes, debido a la variación aleatoria, para conseguir una significación estadística, incluso si no existen en la naturaleza asociaciones reales básicas entre las variables. Cuantas más comparaciones se hagan, más probable es que alguna de ellas sea estadísticamente significativa. Como se dice, «el que busca encuentra».

Este fenómeno se conoce como el problema de las **comparaciones múltiples.** Debido a este problema, la validez de la evidencia de la investigación clínica depende de cómo se enfocaron las preguntas al principio.

Desgraciadamente, cuando se dispone de los resultados de la investigación, no siempre es posible saber cuántas comparaciones se realizaron realmente. A veces, los resultados interesantes se han seleccionado de un gran número de resultados sin interés que ni se

booksmedicos.org

mencionan. El proceso de decidir lo que es y lo que no es importante acerca de una gran cantidad de datos puede, tras ello, provocar una distorsión considerable de la realidad. En la tabla 11-3 se muestra un resumen de cómo se puede producir esta situación engañosa.

Tabla 11-3

Cómo pueden ser erróneas las comparaciones múltiples

1. Realizar muchas comparaciones en el estudio
2. Aplicar pruebas de significación estadística a cada comparación
3. Encontrar algunas comparaciones que sean «interesantes» (estadísticamente significativas)
4. Redactar un artículo sobre uno de estos resultados interesantes
5. No mencionar el contexto la comparación individual (cuántas cuestiones se examinaron y cuál se consideró primaria antes de que los datos fuesen examinados)
6. Elaborar un razonamiento posterior para la verosimilitud del resultado aislado

¿Cómo se pueden tener en cuenta los efectos estadísticos de comparaciones múltiples cuando se interpreta la investigación? Aunque se han propuesto varias formas de adaptar los valores de *P*, probablemente el mejor consejo es ser consciente del problema, y ser prudente al aceptar conclusiones positivas de estudios en los que se realizan comparaciones múltiples. Como afirmó Armitage (16):

> Si extrae los datos suficientemente profundos y con la suficiente frecuencia, encontrará algo extraño. Muchos de estos resultados raros se deberán al azar. No insinúo que la obtención de datos no sea una tarea digna, sino que los descubrimientos que al principio no se postularon entre los principales objetivos del ensayo se deben tratar con mucha precaución.

Un caso especial de comparaciones múltiples ocurre cuando los datos de un ensayo clínico se examinan varias veces conforme se acumulan para garantizar que el ensayo se detiene tan pronto como se obtiene una respuesta, con independencia del tiempo que se calculó que duraría el estudio. Si esto se lleva a cabo, como se hace a menudo por razones éticas, el valor final de *P* suele ajustarse según el número de miradas en los datos.

booksmedicos.org

Es deseable a nivel estadístico mantener la cantidad de miradas al mínimo. En cualquier caso, si los datos resultantes son examinados en varias ocasiones, sería más difícil alcanzar la significación estadística después de tomar en cuenta las múltiples miradas.

Otro caso especial son los estudios de asociación de todo el genoma, en los que más de 500 000 polimorfismos de un solo nucleótido pueden ser examinados en los casos y los controles (17). Una forma habitual de manejar las comparaciones múltiples para dividir el valor habitual de *P* de 0,05 entre el número de comparaciones, requeriría un nivel de significación estadística de 0,00000001 (10^{-7}) para todo el genoma, que sería difícil de alcanzar debido a que los tamaños de muestra en estos estudios se ven limitados y los riesgos relativos generalmente son pequeños. Por esta razón, la confianza en los resultados de los estudios de asociación en todo el genoma depende de la consistencia y la fuerza de las asociaciones en muchos estudios.

ANÁLISIS DE SUBGRUPOS

Es tentador ir más allá de los resultados principales del estudio para examinarlos en subgrupos de pacientes con características que podrían estar relacionadas con el efecto del tratamiento (p. ej., para buscar la modificación del efecto, como se comenta en el capítulo 5). Debido a que las características presentes en el momento de la aleatorización se asignan al azar en los grupos de tratamiento, la consecuencia del análisis en subgrupos es dividir el ensayo como un todo en conjuntos de pequeños ensayos aleatorizados controlados, cada uno con un tamaño de muestra más pequeño.

Ejemplo

En pacientes de alto riesgo, la fibrilación auricular se trata con anticoagulantes antagonistas de la vitamina K para prevenir una enfermedad cerebrovascular. Los investigadores estudiaron la prevención de esta enfermedad en pacientes con fibrilación auricular que no podían tomar antagonistas de la vitamina K. A estos pacientes se les asignó al azar la ingesta del nuevo anticoagulante apixaban o de la aspirina (18). Se presentaron 164 episodios de resultado y el riesgo relativo fue de 0,45 (IC del 95 % de 0,32-0,62) a favor del apixaban. La figura 11-7 muestra el riesgo relativo para enfermedad cerebrovascular en los subgrupos, con las estimaciones puntuales señaladas por los cuadros (con un tamaño proporcional al tamaño del

booksmedicos.org

subgrupo) y los intervalos de confianza. Se informó de un total de 42 subgrupos, 21 para un resultado (enfermedad cerebrovascular y embolia sistémica, que se muestran) y los mismos 21 para otro resultado (hemorragia grave, que no se muestra). Los riesgos relativos para la mayoría de los subgrupos están alrededor del riesgo relativo global para el estudio (que se indica con una línea discontinua vertical), y aunque algunos fueron mayores o menores, ninguno fue estadísticamente significativo. Este análisis sugiere que no se presentó la modificación del efecto, pero esta conclusión estuvo limitada por la baja precisión estadística en los subgrupos. Las comparaciones múltiples, que dan lugar a un resultado falso positivo en los subgrupos, podrían haber sido un problema si hubiera habido un efecto estadísticamente significativo en uno o más subgrupos.

Los análisis por subgrupos informan a los médicos sobre la modificación del efecto para que puedan adaptar el tratamiento de los pacientes individuales, tanto como sea posible, conforme a los resultados del estudio para los pacientes similares a ellos. Sin embargo, el análisis por subgrupos incurre en el riesgo de resultados engañosos debido a la mayor probabilidad de encontrar efectos en un subgrupo particular que no están presentes a largo plazo de manera natural, es decir, encontrar resultados falsos positivos debido a las múltiples comparaciones.

En la práctica, los efectos de las comparaciones múltiples pueden no ser tan extremos cuando los efectos del tratamiento en varios subgrupos no son independientes el uno del otro. A medida que las variables están relacionadas entre sí, en lugar de ser independientes, el riesgo de los hallazgos falsos positivos se reduce. En el ejemplo de la fibrilación auricular y los anticoagulantes, la edad y el antecedente de enfermedad cerebrovascular son componentes de la puntuación CHADS2 (una medida del riesgo de enfermedad cerebrovascular), pero los tres son tratados como subgrupos separados.

Otro riesgo es llegar a una conclusión falsa negativa. Dentro de los subgrupos definidos por cierto tipo de pacientes o tipo específico de resultados hay un menor número de pacientes para todo el estudio, a menudo muy pocos para descartar los hallazgos falsos negativos. Los estudios están, después de todo, diseñados para tener suficientes pacientes para responder a la pregunta principal de investigación con suficiente poder estadístico. Normalmente no están diseñados para tener un poder estadístico suficiente entre los subgrupos, donde el número de pacientes y los episodios de resultado son menores. En la tabla 11-4 se muestran algunas pautas para decidir si el resultado en un subgrupo es real.

booksmedicos.org

Resultados múltiples

Otra versión de las múltiples miradas a los datos es reportar múltiples resultados: diferentes manifestaciones de la efectividad, resultados intermedios y los daños. En general, esto se maneja etiquetando uno de los resultados como el principal y al resto como secundarios, para después ser más cauteloso sobre las conclusiones de los resultados secundarios. Al igual que en los subgrupos, los resultados tienden a estar relacionados entre sí desde una perspectiva biológica (y como una consecuencia estadística), como en el caso del ejemplo anterior, donde la enfermedad cerebrovascular y la embolia sistémica son diferentes manifestaciones del mismo fenómeno clínico.

booksmedicos.org

Característica	N.º de pacientes	Aspirina *(n.º de episodios)*	Apixaban *(%/año)*	Riesgo relativo con apixaban (IC del 95 %)
General	5 599	113 (3,7)	51 (1,6)	
Edad				
< 65 años	1 714	19 (2,0)	7 (0,7)	
65 a < 75 años	1 987	28 (2,7)	24 (2,0)	
≥ 75 años	1 897	66 (6,1)	20 (2,0)	
Género				
Mujer	2 321	64 (4,9)	25 (1,9)	
Hombre	3 277	49 (2,7)	26 (1,4)	
TFG estimada				
< 50 ml/min	1 198	36 (5,8)	16 (2,5)	
50 a < 80 ml/min	2 374	59 (4,5)	22 (1,7)	
≥ 80 ml/min	2 021	18 (1,6)	13 (1,1)	
Puntuación CHADS2				
0-1	2 026	18 (1,6)	10 (0,9)	
2	1 999	40 (3,7)	25 (2,1)	
≥ 3	1 570	55 (6,3)	16 (1,9)	
Antecedente de ECV o ataque isquémico transitorio				
No	4 835	80 (3,0)	41 (1,5)	
Sí	764	33 (8,3)	10 (2,5)	
Dosis de aspirina para el estudio				
< 162 mg diarios	3 602	85 (4,3)	39 (1,9)	
≥ 162 mg diarios	1 978	27 (2,4)	12 (1,1)	
Uso previo de antagonistas de vitamina K				
Sí	2 216	52 (4,2)	17 (1,4)	
No	3 383	61 (3,3)	34 (1,8)	
El paciente se negó a usar antagonistas de vitamina K				
No	3 506	73 (3,8)	35 (1,8)	
Sí	2 092	40 (3,4)	16 (1,45)	
Insuficiencia cardíaca				
No	3 428	66 (3,6)	28 (1,5)	
Sí	2 171	45 (3,8)	23 (1,8)	

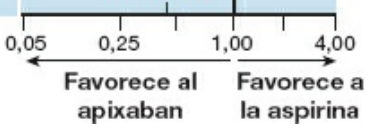


Figura 11-7 ■ Análisis de subgrupos de un ensayo aleatorizado controlado sobre la efectividad del apixaban frente a la aspirina en la enfermedad cardiovascular (ECV) y la embolia sistémica en pacientes con fibrilación auricular. La tasa de filtración glomerular (TFG) es una medida de la función renal. La puntuación CHADS2 es una regla de predicción para el riesgo de embolia en pacientes con fibrilación auricular. (Redibujado con autorización de Connolly SJ,

booksmedicos.org

Eikelboom J, Joyner C, et al. Apixaban in patients with atrial fibrillation. N Engl J Med 2011;364:806-817.)

Tabla 11-4

Directrices para decidir si son reales las diferencias aparentes de los efectos entre subgrupos[a]
A partir del propio estudio: • ¿Es clínicamente importante la magnitud de la diferencia observada? • ¿Qué probabilidad hay de que el efecto haya surgido por azar, teniendo en cuenta: el número de subgrupos examinados? la magnitud del valor *P* • ¿Se formuló la hipótesis del efecto observado antes de descubrirse este (o la justificación del efecto se produjo tras su descubrimiento)? • ¿Era una entre un pequeño número de hipótesis?
A partir de otra información: • ¿La diferencia surgió de las comparaciones en los estudios, en lugar de entre ellos? • ¿Se ha observado el efecto en otros estudios? • ¿Existen pruebas indirectas que respalden la existencia del efecto?

[a]Adaptado de Oxman AD, Guyatt GH. A consumer's guide to subgroup analysis. Ann Intern Med 1992;116:78-84.

MÉTODOS MULTIVARIADOS

La mayor parte de los fenómenos clínicos son consecuencia de muchas variables que actúan conjuntamente de manera compleja. Por ejemplo, la cardiopatía coronaria es el resultado de anomalías lipídicas, hipertensión, tabaquismo, antecedentes familiares, diabetes, dieta, ejercicio, inflamación, coagulopatía y, quizá, la personalidad. Es conveniente intentar comprender estas relaciones examinando, en primer lugar, la clasificación, relativamente sencilla, de los datos, como los análisis estratificados que estudian si el efecto de una variable cambia por la presencia o la ausencia de una o más variables. Cuando los datos se disponen de esta manera, es relativamente fácil comprenderlos.

booksmedicos.org

Sin embargo, como señalamos en el capítulo 7, con este método sólo suele ser posible justificar algunas variables, porque no hay los suficientes pacientes con cada combinación de características como para poder calcular tasas estables. Por ejemplo, si se estudian 120 pacientes, 60 en cada grupo de tratamiento, y sólo se considera una variable dicotómica adicional, sólo habrá, como máximo, unos 15 pacientes de cada subgrupo; si los pacientes se distribuyeron de modo desigual entre los subgrupos, en algunos habrá incluso menos.

Por ello, es necesario, además de las tablas con múltiples grupos, un modo de examinar los efectos de varias variables al mismo tiempo, lo que se logra con el **modelo multivariado** (o desarrollo de una expresión matemática de los efectos de muchas variables tomadas en conjunto). Es «multivariado» porque examina los efectos de múltiples variables simultáneamente. Es un «modelo» porque es una construcción matemática, calculada a partir de los datos basados en hipótesis sobre las características de los datos (p. ej., que todas las variables se distribuyen de manera normal, o que todas tienen la misma varianza).

En la investigación clínica, los modelos matemáticos se emplean de dos formas generales. Una forma es estudiar el efecto independiente de una variable sobre el resultado, mientras se consideran los efectos de otras variables que pueden confundir o modificar esta relación (como se comenta en el cap. 5). La segunda forma es predecir un episodio clínico mediante el cálculo del efecto combinado de varias variables que actúan simultáneamente (el concepto se introduce según las reglas de predicción clínica del cap. 7).

La estructura básica de un modelo multivariado es la siguiente:

$$\text{Variable de respuesta} = \text{constante } (\beta_1 \times \text{variable}_1) + (\beta_2 \times \text{variable}_2) + \ldots,$$

donde β_1, β_2 ... son coeficientes determinados por los datos, y las variable_1, variable_2,... son las variables que se pueden relacionar con el resultado. Las mejores estimaciones de los coeficientes se determinan matemáticamente y dependen de la potencia y capacidad de calcular de las computadoras modernas.

El modelado se realiza de muchas maneras diferentes, pero algunos elementos de este proceso son básicos.

booksmedicos.org

1. Identificación de todas las variables que se pueden relacionar con el resultado de interés, ya sea como factores de confusión o modificadores del efecto. De una forma práctica, podría no ser posible medir en realidad todas ellas, y las variables que faltan deberían mencionarse explícitamente como una limitación.
2. Si hay relativamente pocos episodios de resultado, debe reducirse el número de variables consideradas en el modelo a una cantidad manejable, generalmente sólo unas cuantas. Esto se suele realizar mediante la selección de variables que, cuando se eligen una a una, están más estrechamente relacionadas con el resultado. Si, en esta fase, se utiliza un criterio estadístico, es habitual cometer un error al incluir variables, por ejemplo, al escoger todas las variables que muestran relación con el resultado de interés en un valor de corte de $P < 0{,}010$. Al hacer la selección, también se considera la evidencia para la importancia biológica de la variable.
3. Los modelos, como otras pruebas estadísticas, se basan en las hipótesis sobre la estructura de los datos. Los investigadores deben comprobar si estas hipótesis se cumplen en sus datos particulares.
4. En cuanto a los modelos reales, hay muchas clases y varias estrategias que se pueden seguir dentro de los modelos. Todas las variables, la exposición, el resultado y las covariables, se introducen en el modelo con un orden determinado por la pregunta de investigación. Por ejemplo, si algunas se han de controlar en un análisis causal, se introducen primero en el modelo, seguidas de la variable de interés principal. A continuación, el modelo identificará el efecto independiente de la variable principal. Por otro lado, si el investigador científico quiere hacer una predicción basada en diversas variables, la fuerza relativa de su asociación con la variable de resultado se determina mediante el modelo.

Ejemplo

El cáncer gástrico es la segunda causa principal de muerte por cáncer en el mundo. Investigadores europeos analizaron los datos de una cohorte reclutada en 10 países de este continente para determinar si el alcohol era un factor de riesgo independiente para el cáncer gástrico (19). Identificaron nueve variables, que eran factores de riesgo conocidos, como posibles factores de confusión entre la exposición principal (consumo de alcohol) y la enfermedad (cáncer gástrico): edad, lugar del estudio, sexo, actividad física, educación,

booksmedicos.org

tabaquismo, dieta, índice de masa corporal y en un subgrupo de pacientes, la evidencia serológica de la infección por *Helicobacter pylori*. Como una limitación, mencionaron que ellos habrían incluido el consumo de sal, pero no tenían acceso a los datos sobre su ingesta. El modelado se basó en el modelo de riesgos proporcionales de Cox, por lo que verificaron que la hipótesis subyacente, que el riesgo no variaría con el tiempo, se cumplía. Después ajustar el resto de las variables, el consumo excesivo de alcohol, pero no el ligero, se asoció con cáncer gástrico (riesgo relativo 1,65, IC del 95 % 1,06-2,58); la cerveza estaba asociada, no así el vino y el licor.

Algunos tipos de modelos muy utilizados son la regresión logística (para variables de resultados dicotómicos como los que se producen en estudios de casos y controles) y los modelos de riesgo proporcional de Cox (para los estudios de tiempo para el episodio o evento).

El modelo multivariado es una estrategia esencial para hacer frente a los efectos conjuntos de múltiples variables. No existe otra forma alguna de ajustar ni de incluir variables al mismo tiempo. Sin embargo, esta ventaja tiene un precio. Los modelos tienden a ser cajas negras, y es difícil «acceder al interior» y comprender su funcionamiento. Su validez se basa en suposiciones sobre los datos que puede que no se cumplan. Presentan dificultades para reconocer diferentes efectos en diferentes subgrupos de pacientes. Una exposición variable puede estar muy relacionada con el resultado, pero no aparece en el modelo porque se produce en raras ocasiones, y hay poca información directa sobre la potencia estadística del modelo para esa variable. Por último, los resultados del modelo se ven fácilmente afectados por las peculiaridades de los datos, los resultados de la variación aleatoria en las características de los pacientes de una muestra a otra. Por ejemplo, se ha demostrado que un modelo a menudo identificaba un conjunto diferente de variables predictivas y producía un orden diferente de variables sobre diferentes muestras aleatorias del mismo grupo de datos (20).

Por estos motivos, los modelos en sí no se pueden tomar como patrones de validez, y se deben validar de forma independiente. Normalmente, esto se lleva a cabo observando si el modelo predice o no lo que se encuentra en otra muestra independiente de pacientes. Los resultados del primer modelo se consideran una hipótesis que se debe comprobar por nuevos datos. Si la variación aleatoria es básicamente la responsable de los resultados del primer modelo, no es probable que los mismos efectos aleatorios también se produzcan en el conjunto que se valida. Otra prueba para la validación de un modelo es su verosimilitud biológica y su consistencia con análisis

booksmedicos.org

de datos más transparentes y sencillos, como los análisis estratificados.

RAZONAMIENTO BAYESIANO

Un enfoque totalmente diferente de la información que ha aportado un estudio se basa en la inferencia bayesiana. Se comentó este enfoque en el capítulo 8, donde se aplicó al caso específico de la prueba diagnóstica.

La inferencia bayesiana empieza con una creencia previa sobre la respuesta al tema de la investigación, análoga a la probabilidad preprueba de una prueba diagnóstica, que se basa en todo lo que se conoce sobre la respuesta hasta el momento en que un nuevo estudio ha contribuido a una nueva información. Así pues, cuestiona en qué grado los resultados del nuevo estudio modifican esa convicción.

Algunos aspectos de la inferencia bayesiana son concluyentes. Los estudios individuales no se llevan a cabo con un vacío de información, sino en el contexto de toda la información disponible del momento. El hecho de empezar cada estudio a partir de la hipótesis nula (que no hay ningún efecto) no es real, ya que algo se sabe sobre la respuesta a la cuestión, antes incluso de empezar el estudio. Por otro lado, los resultados de los estudios individuales cambian la creencia en relación con sus fortalezas científicas y la dirección y magnitud de sus resultados. Por ejemplo, si todos los estudios anteriores son negativos y el siguiente, con una fortaleza comparable, es positivo, el efecto todavía es poco probable. Por otro lado, un estudio previo poco consistente puede cambiarse por un único estudio consistente. Finalmente, en este enfoque, no es importante si algunas hipótesis se identifican con antelación y las comparaciones múltiples no son un problema. Antes bien, la convicción previa depende de la verosimilitud de la afirmación, y no de si la afirmación se estableció antes o después de que se iniciara el estudio.

A pesar de que la inferencia bayesiana resulta interesante, hasta el momento ha sido difícil aplicarla, debido al escaso desarrollo de métodos para asignar números a la hipótesis previa y a la información aportada por el estudio. Las dos excepciones son los resúmenes acumulativos sobre la evidencia de la investigación (v. cap. 13) y en las pruebas diagnósticas se produce una excepción, ya que en ellas la «hipótesis» es anterior a la probabilidad, y la nueva información se expresa como un cociente de probabilidades. Sin embargo, la inferencia bayesiana es la base conceptual

booksmedicos.org

para el juicio cualitativo sobre la causa (v. cap.12).

Preguntas de revisión

Lea las siguientes preguntas y marque la respuesta más adecuada.

11-1. Un estudio controlado y aleatorizado del tratamiento trombolítico frente a la angioplastia para el infarto agudo de miocardio no detectó ninguna diferencia en la supervivencia para recibir el alta del hospital. Los investigadores científicos estudiaron si esto también se cumplía para subgrupos de pacientes definidos por la edad, la cantidad de vasos afectados, la fracción de eyección, la comorbilidad y otras características del paciente. ¿Cuál de las siguientes afirmaciones sobre estos análisis de subgrupos *no* es verdadera?

A. Examinar subgrupos aumenta el riesgo de obtener un resultado positivo falso (y se omite significación estadística) en una de las comparaciones.

B. Examinar subgrupos aumenta la posibilidad de obtener un resultado falso negativo en uno de los subgrupos con respecto al resultado principal.

C. Los análisis de subgrupos son prácticas científicas poco consistentes, y no se deberían realizar.

D. La notificación de resultados de los subgrupos facilita a los médicos la adaptación de la información del estudio a los pacientes concretos.

11-2. En un ensayo controlado y aleatorizado de 10 000 pacientes se comparó un nuevo fármaco para la hiperlipidemia con un placebo. Después de 2 años, el colesterol sérico (el resultado principal) fue de 238 mg/dl en el grupo al que se le administraba el fármaco nuevo y 240 mg/dl en el grupo control ($P < 0{,}001$). ¿Cuál de los siguientes describe mejor el significado del valor de *P* en este estudio?

A. No es probable que el sesgo justifique la diferencia observada.

B. La diferencia es clínicamente importante.

booksmedicos.org

C. Habría aparecido, por casualidad, una diferencia tan grande o mayor que la observada en 1 de cada 1 000.

D. Los resultados se pueden generalizar a otros pacientes con hipertensión.

E. La potencia estadística de este estudio fue inadecuada.

11-3. En un ensayo clínico bien diseñado sobre el tratamiento del cáncer ovárico, el índice de remisión al cabo de 1 año es del 30 % en los pacientes a los que se les ofrece un nuevo fármaco y del 20 % para los que se les ofrece placebo. El valor *P* es de 0,4. ¿Cuál de los siguientes describe mejor la interpretación de este resultado?

A. Ambos tratamientos son eficaces.

B. Ningún tratamiento es eficaz.

C. La potencia estadística de este estudio es del 60 %.

D. La mejor estimación de la magnitud del efecto del tratamiento es 0,4.

E. No hay información suficiente para decir si un tratamiento es mejor que el otro.

11-4. En un estudio de cohorte se demostró que la ingesta de vitamina A era un factor de riesgo de sufrir fractura de cadera en las mujeres. El riesgo relativo (quintil mayor frente a quintil menor) era de 1,48, y el IC del 95 % estaba entre 1,05 y 2,07.

¿Cuál de los siguientes describe mejor el significado de este IC?

A. La asociación no es estadísticamente significativa siendo el nivel de $P < 0,05$.

B. Se estableció una estrecha relación entre la ingesta de vitamina A y la fractura de cadera.

C. La potencia estadística de este estudio es del 95 %.

D. Hay un 95 % de posibilidades de que un rango de riesgos relativos de tan sólo 1,05 y de hasta 2,07 incluya el riesgo verdadero.

E. El sesgo es una explicación improbable para este resultado.

11-5. ¿Cuál de las siguientes es la mejor razón para considerar que una *P* ≤ 0,05 es «estadísticamente significativa»?

booksmedicos.org

A. Descarta definitivamente una conclusión falsa positiva.
B. Es una regla de oro elegida arbitrariamente, pero útil.
C. Descarta un error de tipo II.
D. Es una forma de establecer un tamaño del efecto clínicamente importante.
E. Los valores de P mayores o menores no proporcionan información útil.

11-6. ¿Cuál de las siguientes es la mayor ventaja del modelo multivariado?

A. Los modelos pueden controlar muchas variables de manera simultánea.
B. Los modelos no dependen de las hipótesis de los datos.
C. Es un enfoque estandarizado y reproducible sobre el modelado.
D. Los modelos hacen que el análisis estratificado sea innecesario.
E. Los modelos pueden controlar la confusión en los estudios aleatorizados controlados grandes.

11-7. Un ensayo asigna al azar a 10 000 pacientes a dos grupos de tratamiento de tamaño similar, a uno se le ofrece la prevención farmacológica y a otro el tratamiento habitual. ¿Con qué frecuencia debe ocurrir un efecto adverso de la prevención farmacológica para que el estudio tenga una buena probabilidad de observar al menos uno de estos efectos secundarios?

A. 1/5 000.
B. 1/10 000.
C. 1/15 000.
D. 1/20 000.
E. 1/30 000.

11-8. ¿Cuál de las siguientes variables está menos relacionada con el poder estadístico de un estudio de resultado dicotómico?

A. Tamaño del efecto.
B. Error tipo I.
C. Tasa de episodios de resultado en el grupo control.

booksmedicos.org

D. Error tipo II.

E. La prueba estadística utilizada.

11-9. ¿Cuál de las siguientes afirmaciones caracteriza mejor a la aplicación del razonamiento bayesiano en un ensayo clínico?

A. La creencia previa en la eficacia comparativa del tratamiento se rige por el equilibrio.

B. Los resultados de cada estudio nuevo cambian la creencia sobre el efecto del tratamiento que se tenía antes de este.

C. La inferencia bayesiana es una forma alternativa de calcular el valor de *P*.

D. El razonamiento bayesiano se basa en la hipótesis nula, al igual que la estadística inferencial.

E. El razonamiento bayesiano depende de una hipótesis bien definida antes de comenzar el estudio.

11-10. En un ensayo aleatorizado sobre la disminución intensiva de glucosa en la diabetes tipo 2, la tasa de muerte fue mayor en los pacientes tratados de manera estricta: el riesgo relativo fue de 1,22 (IC del 95 % de 1,01-1,46). ¿Cuál de las siguientes afirmaciones acerca de este estudio *no* es verdadera?

A. Los resultados son consistentes con casi ningún efecto.

B. La mejor estimación del efecto del tratamiento es un riesgo relativo de 1,22.

C. Si se calculara el valor de *P*, los resultados serían estadísticamente significativos en un nivel de 0,05.

D. Un valor de *P* podría proporcionar tanta información como el IC.

E. Los resultados son consistentes con unas tasas de mortalidad del 46 % más altas en los pacientes con tratamiento estricto.

Encontrará las respuestas en el apéndice A.

BIBLIOGRAFÍA

1. Fisher R in Proceedings of the Society for Psychical Research, 1929, quoted in Salsburg D. The Lady Tasting Tea. New York: Henry Holt and Co; 2001.
2. Johnson AF. Beneath the technological fix: outliers and probability statements. J Chronic Dis

booksmedicos.org

1985;38:957–961.
3. Courtney C, Farrell D, Gray R, et al for the AD2000 Collaborative Group. Long-term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): randomized double-blind trial. Lancet 2004:363:2105–2115.
4. Bernard SA, Gray TW, Buist MD, et al. Treatment of comatose survivors of out-of-hospital cardiac arrest with induced hypothermia. N Engl J Med 2002;346:557–563.
5. The AIM-HIGH Investigators. Niacin in patients with low HDL cholesterol levels receiving intensive statin therapy. N Engl J Med 2011;365:2255–2267.
6. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. Br J Cancer 1976;34: 585–612.
7. Lind J. A treatise on scurvy. Edinburgh; Sands, Murray and Cochran, 1753 quoted by Thomas DP. J Royal Society Med 1997;80:50–54.
8. Weinstein SJ, Yu K, Horst RL, et al. Serum 25-hydroxyvitamin D and risks of colon and rectal cancer in Finnish men. Am J Epidemiol 2011;173:499–508.
9. Rossouw JE, Anderson GL, Prentice RL, et al. for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principle results from the Women's Health Initiative randomized controlled trial. JAMA 2002;288:321–333.
10. Braitman LE. Confidence intervals assess both clinical significance and statistical significance. Ann Intern Med 1991;114: 515–517.
11. Mai PL, Wideroff L, Greene MH, et al. Prevalence of family history of breast, colorectal, prostate, and lung cancer in a population-based study. Public Health Genomics 2010;13:495–503.
12. Venge P, Johnson N, Lindahl B, et al. Normal plasma levels of cardiac troponin I measured by the high-sensitivity cardiac troponin I access prototype assay and the impact on the diagnosis of myocardial ischemia. J Am Coll Cardiol 2009;54: 1165–1172.
13. McCormack K, Scott N, Go PMNYH, et al. Laparoscopic techniques versus open techniques for inguinal hernia repair. Cochrane Database Syst Rev 2003;(1):CD001785.
14. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. Ann Intern Med 1994;121: 200–206.
15. Sackett DL, Haynes RB, Gent M, et al. Compliance. In: Inman WHW, ed. Monitoring for Drug Safety. Lancaster, UK: MTP Press; 1980.
16. Armitage P. Importance of prognostic factors in the analysis of data from clinical trials. Control Clin Trials 1981;1:347–353.
17. Hunter DJ, Kraft P. Drinking from the fire hose—statistical issues in genomewide association studies. N Engl J Med 2007;357:436–439.
18. Connolly SJ, Eikelboom J, Joyner C, et al. Apixaban in patients with atrial fibrillation. N Engl J Med 2011;364: 806–817.
19. Duell EJ, Travier N, Lujan-Barroso L, et al. Alcohol consumption and gastric cancer risk in European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. Am J Clin Nutr 2011;94:1266–1275.
20. Diamond GA. Future imperfect: the limitations of clinical prediction models and the limits of clinical prediction. J Am Coll Cardiol 1989;14:12A–22A.

booksmedicos.org

¿En qué circunstancias podemos pasar de una asociación observada a un veredicto de causalidad? ¿En qué nos debemos basar para proceder de este modo?

—Sir Austin Bradford Hill
1965

PALABRAS CLAVE

Red de causalidad
Estudios de riesgo agregado
Estudios ecológicos
Falacia ecológica
Estudios de series temporales
Estudio de múltiples series temporales
Análisis de decisión
Análisis de costoefectividad
Análisis de costobeneficio

Este libro ha tratado sobre tres tipos de información clínicamente relevante. Una es la *descripción*, una simple afirmación sobre la frecuencia con la que ocurren las cosas, resumida por mediciones como la incidencia y la prevalencia, así como (en el caso del rendimiento de una prueba diagnóstica) la sensibilidad, la especificidad, el valor predictivo y la razón de verosimilitud. Otra es la *predicción*, la evidencia de que ciertos resultados suelen producirse tras las exposiciones, sin importar si estas son factores de riesgo independientes, por no decir causas. La tercera es directa o

booksmedicos.org

implícitamente sobre la *causa* y el *efecto*. ¿Un factor de riesgo es una causa independiente de la enfermedad? ¿Un factor pronóstico causa un resultado diferente si todo lo demás es igual? Este capítulo trata sobre la causa con gran profundidad.

Otro término para el estudio del origen de una enfermedad es «etiología», que ahora se usa frecuentemente como sinónimo de causa, como en la frase «¿Cuál es la *etiología* de esta enfermedad? En la medida en que la causa de la enfermedad se desconoce, se dice que la enfermedad es «idiopática» o de «etiología desconocida».

Desde hace mucho tiempo existe la tendencia de juzgar la legitimidad de una afirmación causal si tiene sentido de acuerdo con las creencias de la época, como muestran los siguientes ejemplos históricos.

Ejemplo

En 1843, Oliver Wendell Holmes (en aquella época profesor de anatomía y fisiología, y más tarde decano de la Harvard Medical School) publicó un estudio que relacionaba los hábitos del lavado de manos de los obstetras con la fiebre puerperal, una enfermedad generalmente mortal después del parto. (Ahora se sabe que la fiebre puerperal está causada por una infección bacteriana). Las observaciones de Holmes lo llevaron a concluir que: «la afección conocida como fiebre puerperal es contagiosa, ya que los médicos y enfermeras suelen transmitirla de unos pacientes a otros». (1)

La respuesta a la afirmación de Holmes fue que los hallazgos no tenían sentido. «Prefiero atribuirlos (los casos de sepsis puerperal) a la casualidad, o a la Providencia, de la que puedo forjarme una idea, y no al contagio del que no puedo hacerme una idea clara, por lo menos respecto a esta enfermedad en concreto», escribió el Dr. Charles D. Meigs, profesor de obstetricia y enfermedades de la mujer e infantiles en la Jefferson Medical College. Por esos tiempos, un médico húngaro, Ignaz Semmelweis, demostró que desinfectar las manos de los médicos reducía las tasas de fiebre puerperal, y sus estudios también se desecharon porque sus hallazgos no tenían una explicación que se aceptara a nivel general. Las afirmaciones de Holmes y Semmelweis se hicieron décadas antes de la labor pionera de Louis Pasteur, Robert Koch y Joseph Lister, quienes establecieron la teoría de los microorganismos.

La importancia de que una relación causa-efecto «tenga sentido», generalmente en términos de un mecanismo biológico, aún está incorporada en el pensamiento actual. Por ejemplo, en la década de 1990, los estudios que demostraban que la erradicación de la infección por *Helicobacter pylori* prevenía la enfermedad ulcerosa péptica se miraban con escepticismo porque todos sabían que las úlceras gástricas y duodenales no eran una enfermedad infecciosa. En la actualidad se sabe que la infección por *H. pylori* es una

booksmedicos.org

causa importante de la enfermedad.

En este capítulo revisamos los conceptos de causa en la medicina clínica. Comentamos una variedad más amplia de evidencia, además de la plausibilidad biológica, que refuerzan la posibilidad de que una asociación represente una relación causa-efecto. Asimismo, nos ocupamos brevemente de un tipo de diseño de investigación que hasta ahora no se ha considerado en este libro: estudios en los que la exposición a una posible causa sólo la conocen los grupos, y no cada persona del grupo.

PRINCIPIOS BÁSICOS

Causas simples

En 1882, 40 años después de la confrontación de Holmes y Meigs, Koch expuso los postulados para determinar que un agente infeccioso es la causa de la enfermedad (tabla 12-1). Lo esencial para este enfoque era la idea de que una enfermedad específica tiene una causa y que esta causa concreta produce una enfermedad. Este enfoque le ayudó a identificar por primera vez las bacterias que causan tuberculosis, la difteria, la tifoidea y otras enfermedades infecciosas habituales de su época.

Las ideas de Koch contribuyeron en gran medida al concepto de causa en medicina. Antes de Koch se creía que en la causa de una enfermedad determinada influían muchas bacterias diferentes. La aplicación de sus principios ayudó a establecer el orden cuando todo era caos, y todavía son útiles hoy en día. El hecho de que un microorganismo concreto sea la causa de una determinada enfermedad fue la base para que en 1977 se descubriera que la legionelosis se debe a una bacteria gramnegativa; para que en la década de 1980 se determinara que un retrovirus recientemente identificado causaba el SIDA, y el descubrimiento en 2003 que un coronavirus causó un brote de síndrome de insuficiencia respiratoria aguda severa (SARS) (2).

Tabla 12-1

Postulados de Koch
. El microorganismo debe estar presente en cualquier caso de la enfermedad . El microorganismo debe aislarse y crecer en un cultivo puro . El microorganismo debe causar una enfermedad específica cuando se

booksmedicos.org

inocula en un animal
. El microorganismo se debe recuperar del animal e identificarse

Múltiples causas

Para algunas enfermedades, una causa parece ser tan dominante que la consideramos como *la* causa. Decimos que *Mycobacterium tuberculosis* causa la tuberculosis o que un gen anormal que codifica para el metabolismo del aminoácido fenilalanina causa la fenilcetonuria. Podemos pasar por alto el hecho que la tuberculosis también es causada por el huésped y los factores ambientales, y que la enfermedad fenilcetonuria se presenta porque hay fenilalanina en la dieta.

Sin embargo, con mayor frecuencia, varias causas tienen una contribución más equilibrada en la aparición de una enfermedad, por lo que ninguna destaca. La hipótesis subyacente de los postulados de Koch, una causa para una enfermedad, es muy simplificada.

El tabaquismo causa cáncer de pulmón, enfermedades pulmonares obstructivas crónicas, arteriopatía coronaria y arrugas en la piel. La arteriopatía coronaria tiene múltiples causas, como el tabaquismo, la hipertensión, la hipercolesterolemia, la diabetes, la inflamación y la herencia. Ciertos parásitos específicos pueden causar malaria, pero sólo si el mosquito vector puede reproducirse, se infecta y pica a la gente, y estas no están tomando fármacos contra la malaria o son incapaces de controlar la infección por sí mismas.

Cuando muchos factores actúan en conjunto se ha denominado **«red de causalidad»** (3), que puede comprenderse muy bien en las enfermedades crónicas degenerativas, como la enfermedad cardiovascular y el cáncer. También es la base para las enfermedades infecciosas, en las que es necesaria la presencia de un microorganismo pero no es suficiente para causar la enfermedad. El SIDA no puede aparecer sin la exposición al VIH, pero esta exposición al virus no implica necesariamente que se contraiga la enfermedad. Por ejemplo, la exposición al VIH rara vez provoca la seroconversión después de la punción con una aguja (3/1 000) porque el virus no es tan infeccioso como el de la hepatitis B. De forma similar, no todas las personas expuestas a la tuberculosis –en la época de Koch o actualmente– se infectan.

booksmedicos.org

Cuando más de una causa actúa conjuntamente, el riesgo resultante puede ser superior o inferior al previsto, simplemente combinando los efectos de las causas por separado. Es decir, interactúan –hay una modificación del efecto. La figura 12-1 muestra el riesgo de enfermedad cardiovascular en 10 años en un hombre de 60 años sin antecedentes de enfermedad cardiovascular, de acuerdo con la presencia o ausencia de varios factores de riesgo comunes. El riesgo es mayor que la suma de los efectos de cada factor de riesgo individual. El efecto de un HDL bajo es mayor ante la presencia de un colesterol total elevado, el efecto del tabaquismo es mayor ante la presencia de un colesterol total elevado y un HDL bajo, y así sucesivamente. La consecuencia de la exposición a cada factor de riesgo nuevo se ve afectado por la exposición a los demás, lo que constituye un ejemplo de la modificación del efecto. La edad y el sexo también son factores de riesgo e interactúan con los demás (no se muestran).

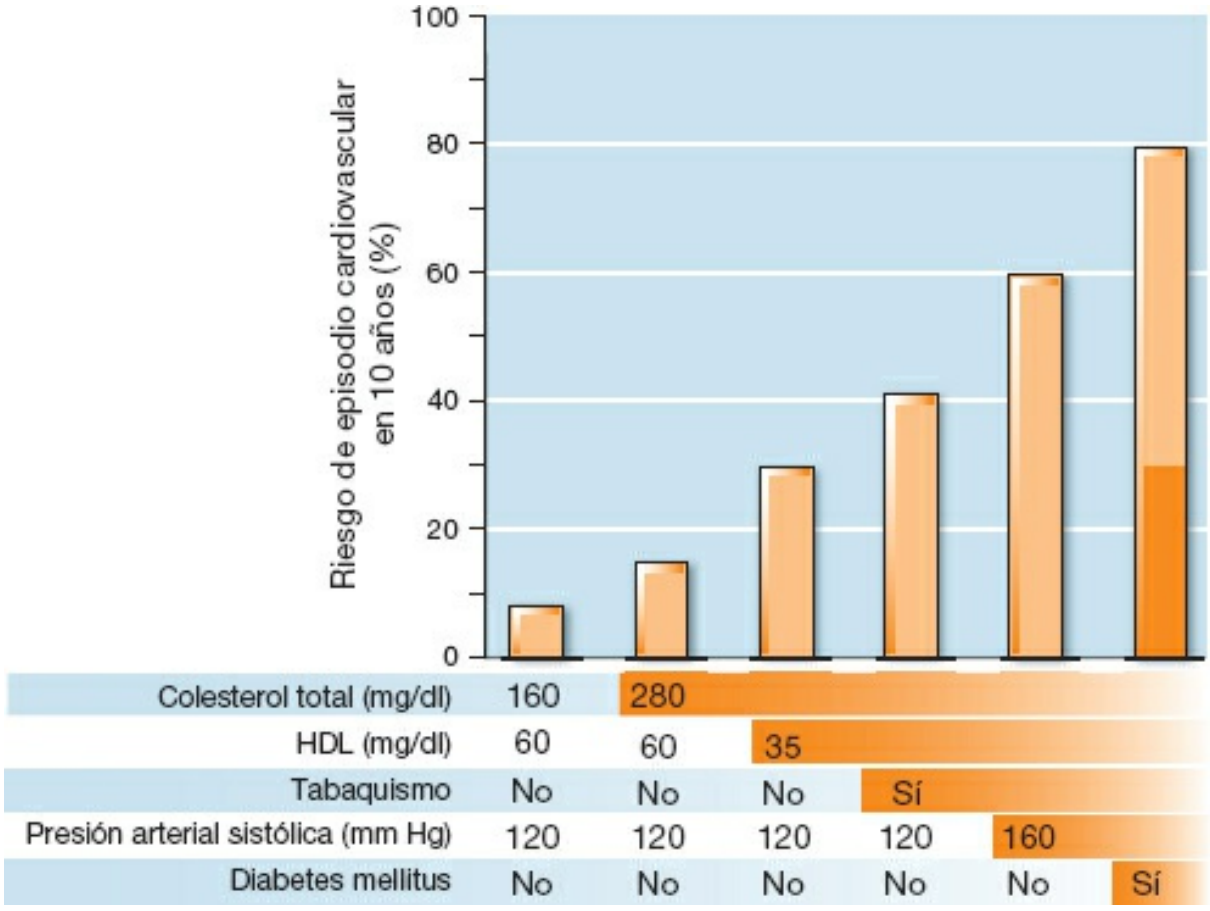


Figura 12-1 ■ La interacción de múltiples factores de riesgo para enfermedad cardiovascular. El riesgo cardiovascular a 10 años (%) para un hombre de 60 años sin factores de riesgo *(barra izquierda)* y con la adición sucesiva de cinco factores de riesgo *(barras derechas)*. Cada factor de riesgo aislado

booksmedicos.org

contribuye relativamente poco (varios puntos porcentuales) al riesgo, mientras que la adición de cada uno de ellos entre sí aumenta el riesgo casi 10 veces, mucho más que la suma de factores de riesgo individuales que actúan de manera independiente, lo que se muestra mediante el área sombreada de la barra de la derecha. (De The Framingham Risk Calculator in UpToDate, Waltham, MA according to formulae in D'Agostino RB Sr, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care. The Framingham Heart Study. Circulation 2008;117[6]:743-753.)

Cuando están presentes múltiples factores causales e interactúan, es posible que el hecho de cambiar sólo una causa, o algunas de ellas, afecte de forma considerable a la salud del paciente. En el ejemplo anterior, si se trata a los pacientes la hipertensión y el aumento de colesterol sérico, se puede disminuir de forma considerable el riesgo de sufrir enfermedades cardiovasculares, incluso si los otros factores de riesgo no cambian.

Con todo, los médicos están más interesados en las causas tratables o reversibles que en las inmutables. Por ejemplo, cuando se trata de la enfermedad cardiovascular, la edad y el sexo no se pueden cambiar, deben de considerarse como son. Por otro lado, el tabaquismo, la presión arterial y el colesterol sérico se pueden modificar. Por lo tanto, aunque los riesgos relacionados con la edad y el sexo son tan grandes como para ser factores de riesgo como el resto, y se tienen en cuenta al estimar el riesgo cardiovascular, no son un objetivo de prevención o tratamiento.

Proximidad de causa-efecto

Cuando los científicos biomédicos estudian la etiología, suelen buscar el mecanismo patogénico subyacente o la vía final habitual de la enfermedad. La anemia drepanocítica es un ejemplo. De forma simplificada, la patogénesis depende de un gen codificado para la hemoglobina anormal que se polimeriza en ambientes con poco oxígeno (los capilares de algunos tejidos), causando eritrocitos deformados, anemia por su destrucción y la oclusión de los vasos, provocando ataques de isquemia con dolor y destrucción del tejido.

La enfermedad también la determinan otras causas más remotas, menos específicas (o factores de riesgo), como la conducta de los pacientes o las características de su entorno. Estos factores pueden tener grandes efectos sobre las tasas de enfermedad. Por ejemplo, una gran proporción de muertes por enfermedades cardiovasculares y cáncer en EE.UU. se debe a factores medioambientales y conductuales, como el tabaquismo, la alimentación y el sedentarismo; el SIDA se transmite sobre todo por conductas sexuales de

booksmedicos.org

riesgo y agujas compartidas, y las muertes por violencia y las lesiones accidentales tienen su origen en las condiciones sociales, el acceso a las armas, conducir intoxicado y el uso del cinturón de seguridad.

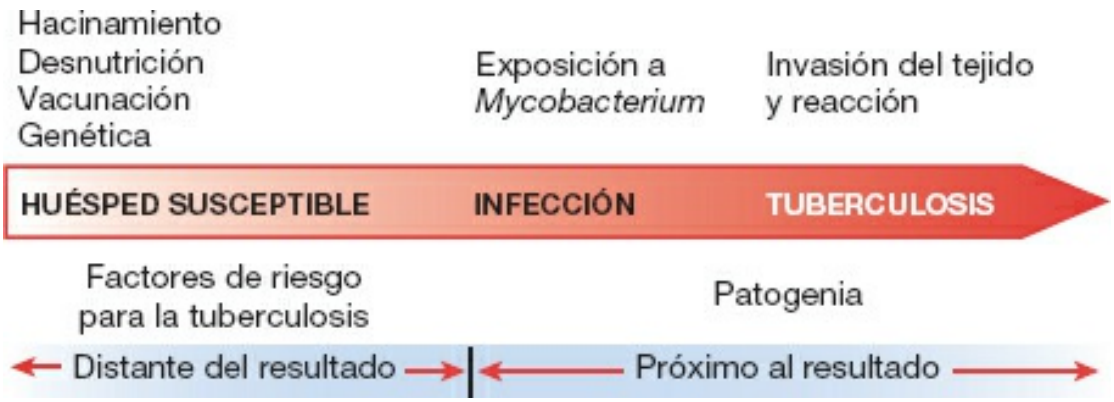


Figura 12-2 ■ Causas lejanas y próximas de tuberculosis.

La figura 12-2 muestra cómo los factores de riesgo y la patogénesis de la tuberculosis –causas distantes y proximales– siguen al continuo de la enfermedad. La exposición a *M. tuberculosis* depende del entorno del huésped: la proximidad a los casos activos. La infección depende de la susceptibilidad del huésped, que puede aumentar por la desnutrición, disminuir por la vacunación y verse alterada por la carga genética. Si la infección progresa a la enfermedad depende de estos y otros factores, como la inmunocompetencia, que puede verse comprometida por la infección por VIH y la edad. Finalmente, la infección activa puede curarse con tratamiento antibiótico.

Los médicos pueden estar tan concentrados en la patogénesis que pueden subestimar la importancia de las causas más remotas de la enfermedad. En el caso de la tuberculosis, las mejoras sociales y económicas que influyen en la susceptibilidad del huésped, como vivir en un sitio sin hacinamiento y una mejor nutrición, parecen haber jugado un papel más relevante que los tratamientos en la disminución de las tasas de tuberculosis en países desarrollados. La figura 12-3 muestra que la tasa de mortalidad por tuberculosis en Inglaterra y Gales disminuyó dramáticamente antes de identificar el bacilo de la tuberculosis y un siglo antes de que los primeros antibióticos efectivos aparecieran en la década de 1950.

La red de causalidad está cambiando continuamente, incluso para las enfermedades antiguas. Entre 1985 y 1992, el número de casos de

booksmedicos.org

tuberculosis en EE.UU., que había estado disminuyendo desde hace un siglo, comenzó a aumentar (fig. 12-4) (4). ¿Por qué pasó esto? Se había producido una afluencia de inmigrantes de los países con tasas altas de tuberculosis. La epidemia de SIDA produjo más personas con un sistema inmunitario debilitado, haciéndolo más susceptible a la infección por *M. tuberculosis*. Cuando se infectan, sus cuerpos permiten la multiplicación masiva de la bacteria, y son más infecciosos. La multiplicación rápida, sobre todo en pacientes que no siguieron los regímenes farmacológicos prescritos, favoreció el desarrollo de cepas resistentes a múltiples fármacos. Las personas que eran más propensas a tener tanto SIDA y tuberculosis –las socialmente desfavorecidas, los consumidores de drogas intravenosas y los prisioneros– estaban desarrollando una enfermedad resistente a múltiples fármacos y exponiendo a otros en la población a una cepa difícil de tratar. La interacción del medio ambiente, la conducta y la biología molecular se combinaron para revertir la tendencia a la baja en la tuberculosis. Para combatir la nueva epidemia de tuberculosis se reconstruyó la infraestructura pública sanitaria. Se iniciaron las pautas terapéuticas con múltiples fármacos (esfuerzos biológicos) como el tratamiento directamente observado para garantizar el cumplimiento (esfuerzos conductuales) y la tasa de tuberculosis comenzó a disminuir de nuevo.

booksmedicos.org

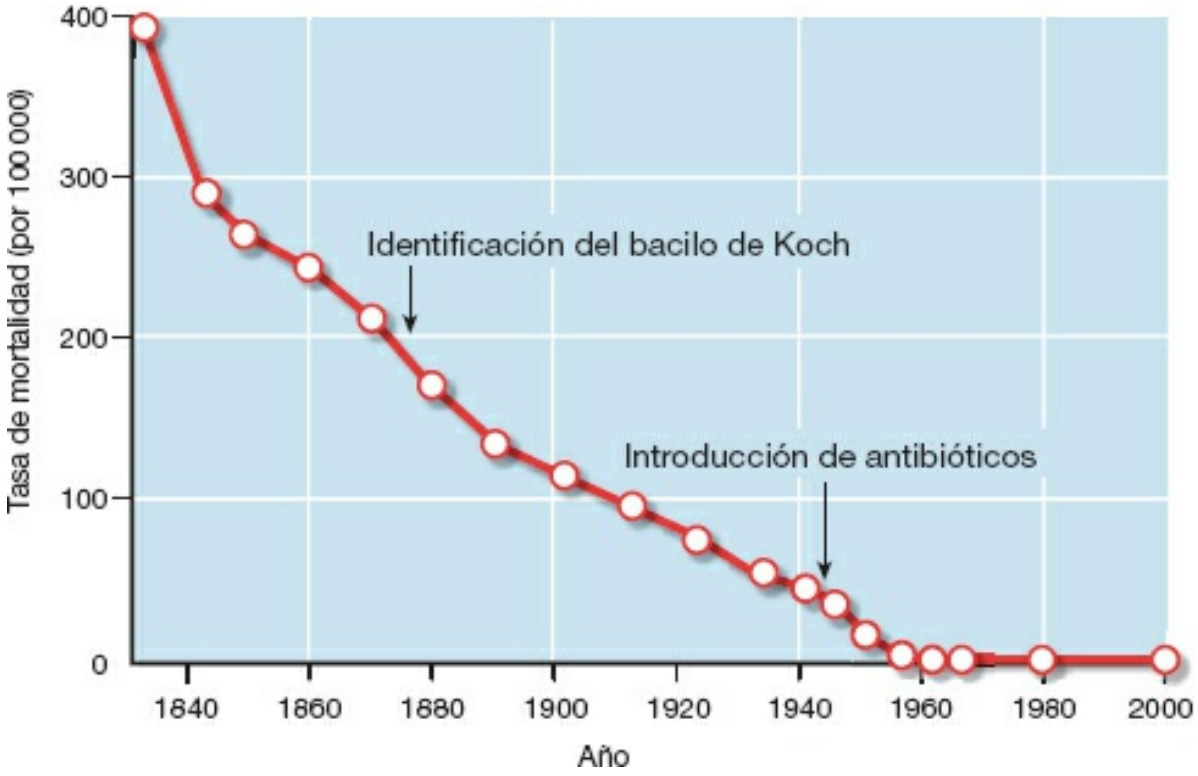


Figura 12-3 ■ Descenso del índice de mortalidad por tuberculosis respiratoria en Inglaterra y Gales durante los últimos 150 años. La mayor parte de la disminución se produjo antes de que estuviera disponible el tratamiento antibiótico. (De McKeown T. The Role of Medicine: Dream, Mirage, or Nemesis. London: Nuffield Provincial Hospital Trust; 1976 y de http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/Tuberculosis/TBUKSurveillanceData/TuberculosisMortality.)

booksmedicos.org

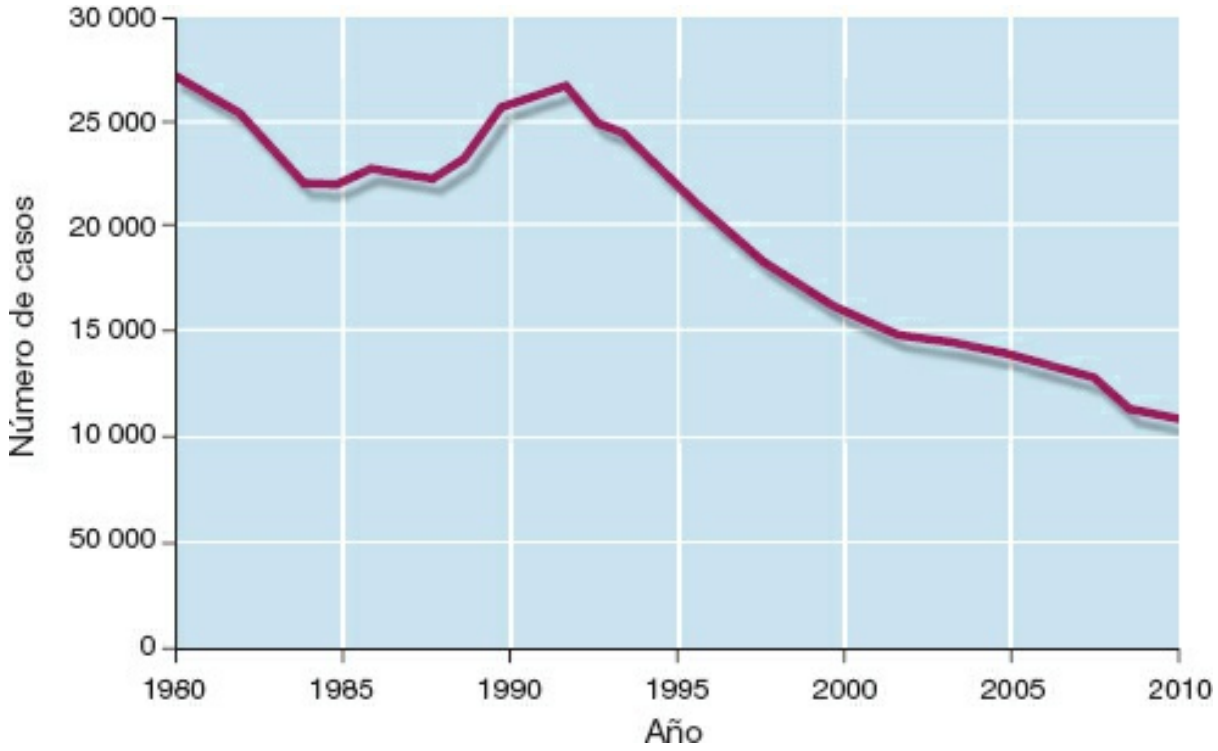


Figura 12-4 ■ Casos de tuberculosis en EE.UU., 1980 hasta 2010. Un descenso de larga evolución se detuvo en 1985. El número de casos alcanzó un máximo en 1992 y luego comenzó a disminuir de nuevo. (Adaptado de Centers for Disease Control and Prevention. Reported tuberculosis in the United States, 2010. Disponible en: http://www.cdc.gov/Features/dsTB-2010Data/. Accessed February 8, 2012.)

EVIDENCIA INDIRECTA DE CAUSALIDAD

En la medicina clínica, no es posible demostrar relaciones causales más allá de toda duda, como se podría hacer con una fórmula matemática. Lo que podemos hacer es aumentar nuestra convicción sobre la relación causa-efecto, mediante la evidencia empírica, hasta el punto de que se haya establecido la causa. En cambio, pueden aumentar las pruebas en contra de una causa hasta que una relación causa-efecto resulte improbable.

La proposición de una relación de causa-efecto se debe examinar de todas las formas posibles. En el resto de este capítulo se discuten algunos enfoques habitualmente utilizados.

Examinando los estudios individuales

A lo largo de este libro se ha discutido un enfoque sobre la evidencia de la causa-efecto: un análisis profundo de los propios estudios. Cuando se ha observado una asociación, se ha establecido una relación causal en la medida

booksmedicos.org

en que la asociación no puede explicarse por el sesgo o el azar. La figura 12-5 resume un enfoque familiar. En primer lugar se buscan los sesgos y cuánto han podido modificar el resultado; después se considera si es poco probable que la asociación se deba al azar. En los estudios observacionales, la confusión siempre es una posibilidad. A pesar de que la confusión se puede controlar de forma integral, según la ciencia, casi nunca es posible descartarla por completo; por lo tanto, la confusión continúa siendo el reto permanente para el razonamiento causal basado en la investigación observacional.

Figura 12-5 ■ Asociación y causa. El sesgo, el azar y la confusión deben excluirse antes de concluir que es probable una asociación causal.

Los ensayos aleatorizados pueden tratar definitivamente la confusión, pero no es posible realizarlos en los estudios de riesgo (p. ej., causas) *per se*. Por ejemplo, no sería ético (y sería poco exitoso) aleatorizar a los no fumadores a fumar cigarrillos para estudiar si fumar causa cáncer de pulmón. Sin embargo, los ensayos aleatorizados controlados pueden contribuir a la inferencia causal en dos situaciones. Una de ellas es cuando el ensayo está enfocado a tratar una posible causa, como el colesterol o la presión sanguínea elevados, y se evita el resultado. Otra es cuando el ensayo se realiza para otro propósito y la intervención causa daños imprevistos. Por ejemplo, el exceso de episodios cardiovasculares en los ensayos aleatorizados de rofecoxib, un inhibidor de la ciclooxigenasa 2, que se administraba por otras razones (p. ej., alivio del dolor), muestra que este fármaco puede ser una causa de episodios

booksmedicos.org

cardiovasculares.

Jerarquía de los diseños de investigación

Se puede establecer una jerarquía de los diferentes diseños de investigación de acuerdo con su fortaleza científica para establecer causalidad (tabla 12-2). En la jerarquía más alta están las revisiones sistemáticas de los ensayos aleatorizados controlados porque tratan definitivamente con la confusión. Estos ensayos están seguidos por los estudios observacionales, con poca diferencia entre los estudios de cohorte y los de casos y controles, en una época donde los análisis de este último tipo de diseño están anidados en las cohortes obtenidas de poblaciones definidas. En la jerarquía más baja aún están los estudios no controlados, el razonamiento biológico y la experiencia personal. Por supuesto, este orden es sólo una guía burda sobre la fuerza de la evidencia. La manera de realizar un estudio individual puede crear la diferencia para aumentar o disminuir su validez, sin importar el tipo de estudio diseñado. Un estudio aleatorizado controlado malo contribuye menos a entender la causa que un estudio de cohorte ejemplar.

Tabla 12-2

Jerarquía de la fortaleza del diseño de investigación
Estudios de riesgo individual
Revisiones sistemáticas: evidencia consistente de múltiples ensayos aleatorizados controlados
Ensayos aleatorizados controlados
Estudios observacionales Estudios de cohortes Estudios de casos y controles, casos y cohorte
Estudios transversales
Series de casos
Experiencia, opinión de expertos

Con esta jerarquía en mente, la fortaleza de la evidencia para la causa y el efecto se juzga a veces de acuerdo con los mejores estudios sobre la pregunta. Los ensayos aleatorizados controlados bien diseñados y bien ejecutados

booksmedicos.org

triunfan sobre los estudios observacionales, los estudios observacionales basados en el conocimiento científico actual, y así sucesivamente. Este es un enfoque muy simplificado de la evidencia pero es un atajo útil.

EL CUERPO DE LA EVIDENCIA A FAVOR Y EN CONTRA DE LA CAUSA

¿Qué aspectos de los hallazgos de investigación apoyan la causa y el efecto cuando sólo están disponibles los estudios observacionales? En 1965, el estadístico inglés Sir Austin Bradford Hill propuso una serie de observaciones que, en conjunto, ayudan a establecer si la relación entre el factor ambiental y la enfermedad es causal o es meramente una asociación (5) (tabla 12-3). Revisamos estos «criterios de Bradford Hill» utilizando principalmente el tabaquismo y el cáncer de pulmón como un ejemplo. En general se cree que el tabaquismo causa cáncer de pulmón, aunque no existen ensayos aleatorizados controlados sobre el tabaquismo o un mecanismo biológico indiscutible.

Tabla 12-3

Evidencia de que una asociación es causa y efecto	
Criterios	**Comentarios**
Temporalidad	Causa precede al efecto
Solidez	Riesgo relativo grande
Dosis-respuesta	Las grandes exposiciones a la causa se relacionan con elevadas tasas de enfermedad
Reversibilidad	La reducción de la exposición va seguida de tasas más bajas de enfermedad
Consistencia	Observada repetidamente por distintas personas, en lugares, circunstancias y momentos diferentes
Plausibilidad biológica	Tiene sentido, según el conocimiento biológico de la época
Especificidad	Una causa produce un efecto
Analogía	La relación causa-efecto ya está establecida para una exposición similar o enfermedad

booksmedicos.org

Adaptado de Bradford Hill AB. The environment and disease: association and causation. Proc R Soc Med 1965;58:295-300.

¿La causa precede al efecto?

Obviamente, la causa debe ocurrir antes de sus efectos. Esto parece evidente, pero este principio puede pasarse por alto cuando se interpretan estudios transversales y de casos y controles, donde tanto la supuesta causa como el efecto se miden en los mismos momentos. El tabaquismo claramente antecede al cáncer de pulmón durante muchas décadas, pero hay otros ejemplos en los que se puede confundir el orden de la causa y el efecto.

Ejemplo

El «latigazo» es la presencia de dolor cervical después de una lesión debida a una flexión/extensión forzada, que suele ocurrir por un accidente automovilístico. Muchas personas con latigazos se recuperan con rapidez, pero otras presentan síntomas crónicos, interponen una demanda y solicitan gratificaciones por discapacidad. Muchos pacientes con latigazos tienen ansiedad y depresión, y los estudios epidemiológicos han vinculado estos síntomas. Se suele creer que la lesión y el dolor asociado causan los síntomas psicológicos. Investigadores noruegos probaron la hipótesis que la causa y el efecto iban en dirección opuesta (6). Analizaron una cohorte en la que se habían medido los síntomas psicológicos antes de las lesiones y la siguieron durante 11 años. Las personas con ansiedad y depresión antes de la lesión eran más propensas de tener un latigazo.

Haber descubierto que lo que se pensaba que era una causa, en realidad obedece a un efecto, es una poderosa evidencia en *contra* de la causa, pero por sí misma la secuencia temporal sólo es una evidencia mínima *a favor de* la causa.

Solidez de la asociación

Una asociación sólida entre una supuesta causa y un efecto, expresada por un gran riesgo absoluto o relativo, es la mejor prueba de una relación causal que una asociación débil. Los sesgos no reconocidos podrían explicar pequeños riesgos relativos, pero es poco probable que produzcan riesgos.

Así pues, la incidencia 20 veces superior del carcinoma pulmonar entre los hombres fumadores en comparación con los no fumadores es una prueba mucho más decisiva de que el tabaquismo provoca este tipo de cáncer, que el

booksmedicos.org

descubrimiento de que el tabaquismo está relacionado con el carcinoma renal, cuyo riesgo relativo es mucho menor (un 1,5). De forma similar, un aumento de 10 o 100 veces el riesgo de carcinoma hepatocelular en pacientes con infección por hepatitis B es una fuerte evidencia que el virus causa cáncer hepático.

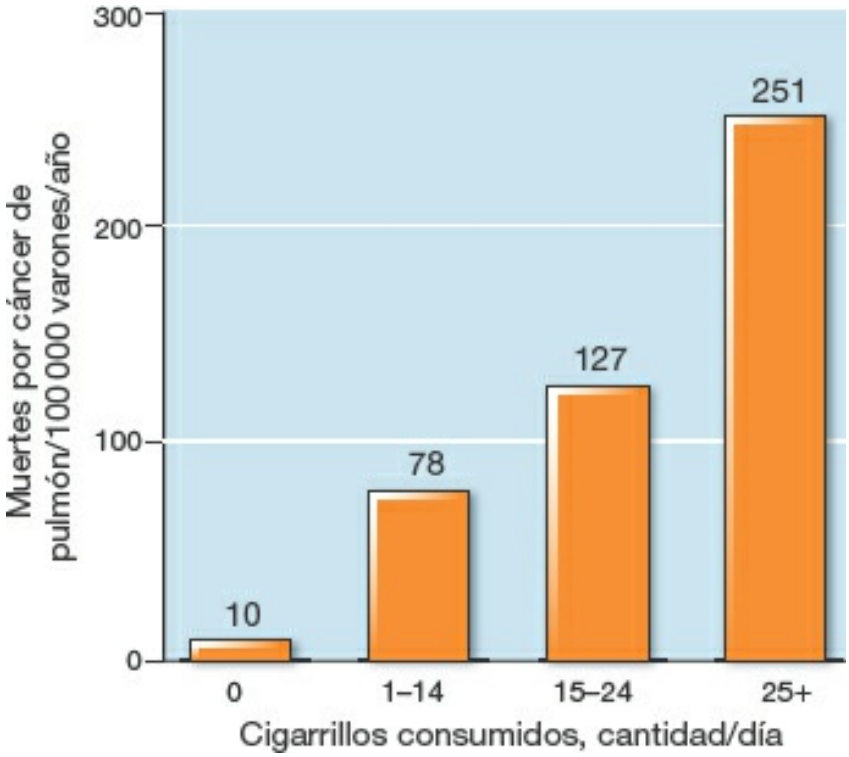


Figura 12-6 ■ Ejemplo de relación dosis-respuesta: muertes de médicos hombres por cáncer de pulmón según la dosis (cantidad) de cigarrillos consumidos. (De in Doll R, Peto R. Mortality in relation to smoking: 20 years' observations on male British doctors. Br Med J 1976;2:1525-1536.)

Relaciones dosis-respuesta

Existe una relación entre dosis-respuesta si, al aumentar la exposición a la supuesta causa, se produce un efecto cada vez mayor. En el caso del tabaquismo, la «dosis» puede ser el número de años de tabaquismo, los paquetes que se consumen al día o los «años-paquete». La figura 12-6 muestra una curva clara de la dosis-respuesta donde se dibujan las tasas de mortalidad por cáncer de pulmón (respuestas) frente al número de cigarrillos fumados (dosis).

El hecho de demostrar la relación dosis-respuesta refuerza la tesis de la causa-efecto, pero la ausencia de esta es una prueba relativamente débil frente

booksmedicos.org

a la causalidad, porque no todas las asociaciones causales presentan una relación del tipo dosis-respuesta dentro del rango observado y porque aún es posible la confusión.

Ejemplo

Tanto la fuerte asociación entre tabaquismo y cáncer de pulmón como la relación entre dosis-respuesta, pueden ser ejemplos de factores de confusión. Según esta afirmación puede existir una variable desconocida que provoque el tabaquismo y aumente el riesgo de sufrir un carcinoma pulmonar. Cuanto más presente esté el factor, más tabaquismo y cáncer de pulmón se detecta; por lo tanto, existe una relación dosis-respuesta. Teóricamente, dicha afirmación es una posible explicación de la asociación entre tabaquismo y cáncer de pulmón, aunque nunca se ha demostrado exactamente cuál es el posible factor de confusión. A falta de ensayos controlados y aleatorizados (que asignarían a las personas el factor de confusión a partes iguales en los grupos de fumadores y no fumadores), es difícil refutar por completo la posibilidad de confusión.

Asociaciones reversibles

Es más probable que un factor sea una causa de enfermedad si su eliminación produce una disminución del riesgo de la afección. La figura 12-7 muestra que cuando se abandona el hábito tabáquico disminuye la probabilidad de presentar un cáncer de pulmón con respecto al número de años desde el abandono.

Las asociaciones reversibles constituyen pruebas concluyentes, pero no infalibles, de una relación causal, pues la confusión podría ser la responsable. Por ejemplo, la figura 12-7 concuerda con la explicación (poco probable) de que los fumadores que desean abandonar el tabaquismo presentan menor cantidad de un factor de confusión no identificado que aquellos que continúan fumando.

Consistencia

Si varios estudios, realizados en momentos diferentes, en distintos ámbitos y con varios tipos de pacientes, llegan a la misma conclusión, se consolidan las pruebas de la relación causal. La causalidad se corrobora especialmente si los estudios utilizan diferentes diseños de investigación, con puntos fuertes y débiles complementarios, que conducen al mismo resultado, porque en los estudios en los que se emplea el mismo diseño se puede cometer el mismo

booksmedicos.org

error. Para demostrar la relación del tabaquismo y el cáncer de pulmón, muchos estudios de cohorte, de casos y controles y de series temporales han demostrado que el aumento del consumo de tabaco conlleva un incremento de la incidencia del cáncer de pulmón, en ambos sexos, en diferentes grupos étnicos y en diversos países.

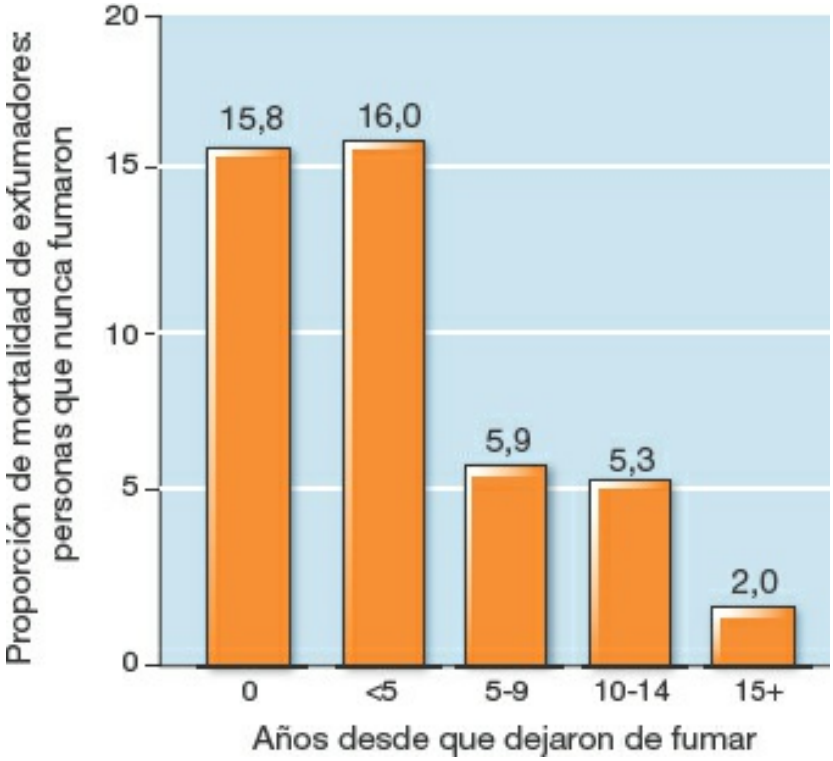


Figura 12-7 ■ Asociación reversible: descenso de la mortalidad por cáncer de pulmón en exfumadores de cigarrillos. Los datos excluyen a las personas que dejaron de fumar tras presentar un cáncer. (De in Doll R, Petro R. Mortality in relation to smoking: 20 years' observations on male British doctors. Br Med J 1976;2:1525-1536.)

Diferentes estudios pueden producir diferentes resultados. La falta de consistencia no necesariamente significa que no existe una relación causal. Los resultados del estudio pueden diferir por las diferencias entre los pacientes, las intervenciones, el seguimiento o la medición del resultado (p. ej., plantear preguntas de investigación un tanto diferentes). Además, los estudios también pueden variar en calidad, y un buen estudio puede contribuir con información más válida que varios estudios deficientes.

Plausibilidad biológica

Como se comentó al principio de este capítulo, cuando se evalúa la causa y

booksmedicos.org

efecto, a menudo se le da un gran peso a la idea de que una posible causa coincide con nuestro conocimiento sobre los mecanismos de la enfermedad, como actualmente se entiende.

Cuando se desconoce por completo cómo puede haber surgido una asociación, se tiende al escepticismo de que la relación sea real, y suele ser así. Tal escepticismo a menudo suele ser útil.

Ejemplo

Al principio de los años 80 se propuso que la sustancia laetrile, extraída de las semillas de chabacano, curaba el cáncer. La comunidad científica no estaba convencida, básicamente porque no había ninguna razón biológica para que un extracto de dichas semillas que no tenía ningún tipo de relación química con los componentes antineoplásicos conocidos fuera eficaz contra las células cancerígenas. No obstante, muchos pacientes con cáncer estaban tomando laetrile, obtenido de fuentes no convencionales, algunas veces en lugar de la quimioterapia convencional. Para resolver esta controversia, algunos investigadores probaron la efectividad del laetrile en un ensayo aleatorizado controlado. El ensayo demostró que la sustancia realmente no tenía ninguna actividad contra los carcinomas estudiados (7).

De estar presente, la plausibilidad biológica refuerza los argumentos de causalidad, pero la ausencia de plausibilidad biológica puede sólo reflejar las limitaciones de la comprensión sobre la biología de la enfermedad, en lugar de la falta de asociación causal.

Especificidad

La especificidad (una causa, un efecto) se suele detectar con mayor frecuencia en enfermedades infecciosas graves (p. ej., poliomielitis y tétanos) y en enfermedades genéticas (p. ej., poliposis adenomatosa familiar u ocronosis), aunque los efectos genéticos suelen ser modificados por interacciones gen-gen y gen-entorno. Como se mencionó anteriormente en este capítulo, las enfermedades crónicas degenerativas a menudo tienen muchas causas para el mismo efecto y muchos efectos para la misma causa. El tabaquismo, el amianto y la radiación causan cáncer de pulmón. El tabaquismo no sólo causa cáncer de pulmón sino también bronquitis, enfermedad cardiovascular, enfermedad periodontal y piel arrugada, por mencionar algunas. Por lo tanto, la especificidad es una evidencia sólida para la causa y el efecto, pero la ausencia de especificidad es una evidencia débil

booksmedicos.org

contra ella.

Analogía

El debate sobre la relación causa-efecto se refuerza si existen ejemplos de causas bien establecidas análogas a la causa en cuestión. Por lo tanto, el caso de que el tabaquismo causa cáncer de pulmón se ve reforzado por las observaciones de que otras sustancias ambientales como el asbesto, el arsénico y el uranio también lo causan.

En cierto sentido, el aplicar los criterios de Bradford Hill a la causa, es un ejemplo de razonamiento bayesiano. Por ejemplo, la creencia en la causalidad basada en la fortaleza de la asociación y la dosis-respuesta se modifican (aumenta o se reduce) por la evidencia concerniente a la plausibilidad biológica o a la especificidad, y cada uno de los criterios contribuyen en mayor o menor medida con la creencia general de que una asociación es causal. La diferencia principal entre el enfoque bayesiano y la prueba diagnóstica es que varias líneas de evidencia para la causa (dosis-respuesta, reversibilidad, consistencia, etc.) se están formando simultáneamente en diversas disciplinas científicas en lugar de formarse en serie mediante la investigación clínica.

ESTUDIOS DE RIESGO AGREGADO

Hasta ahora, hemos comentado los estudios en los que conocemos la exposición y la enfermedad en cada miembro del estudio. Para completar la gama de diseños de investigación, ahora consideraremos un tipo diferente de estudios, denominados **estudios de riesgo agregado,** donde la exposición al factor de riesgo se caracteriza por la exposición promedio del grupo al que pertenecen los individuos. Otro término es **estudios ecológicos,** porque las personas se clasifican por el nivel general de exposición en su entorno, que puede o no corresponder con su exposición individual. Los ejemplos son estudios epidemiológicos que relacionan el consumo de vino de un país con las tasas de mortalidad cardiovascular o bien los estudios sobre las tasas regionales de cáncer regional o defectos congénitos en relación con las exposiciones regionales, como los derrames de productos químicos.

El problema principal con los estudios que simplemente correlacionan la exposición promedio con la tasa promedio de enfermedad en los grupos es la posibilidad de una **falacia ecológica,** en el que las personas afectadas de un

booksmedicos.org

grupo puede que no hayan estado expuestas al factor de riesgo al que se expone el grupo en general. Además, la exposición puede no ser la única característica que diferencie a las personas del grupo expuesto de las del no expuesto, es decir, que puede haber factores de confusión. Por lo tanto, los estudios de riesgo añadido como este son muy útiles para formular hipótesis, que después deben ser probadas mediante una investigación más rigurosa.

La evidencia derivada de los estudios de riesgo añadido puede fortalecerse cuando las observaciones se realizan durante el período que agrupa a la exposición, e incluso fortalecerse aún más si las observaciones se encuentran en más de un sitio y momento.

En un **estudio de series temporales,** las tasas de enfermedad se miden en diferentes momentos del tiempo, tanto antes como después de introducir la supuesta causa. Entonces es posible comprobar si una tendencia en la tasa de enfermedad cambia con el tiempo en relación con el momento de la exposición. Si los cambios de la supuesta causa van seguidos de cambios del supuesto efecto, y no en otro momento, es menos probable que la asociación sea falsa. Una ventaja de los análisis de series temporales es que pueden distinguir los cambios que se producen con el tiempo (tendencia a largo plazo) y los efectos de la intervención.

Ejemplo

Las infecciones por *Staphylococcus aureus* resistente a meticilina (SARM) asociadas con la atención médica son una causa principal de morbilidad, mortalidad y costos en los hospitales de cuidados intensivos. La U.S. Veterans Affairs (VA) inició un programa nacional para prevenir estas infecciones en los hospitales de VA (8). La VA implementó un «paquete de SARM» que comprende un conjunto de intervenciones que incluyen el lavado de manos, la revisión nasal universal para el SARM, las precauciones universales con los pacientes colonizados o infectados por SARM y los cambios en las responsabilidades institucionales para prevenir la infección. Los datos sobre las tasas de infección por SARM se obtuvieron 2 años antes de su aplicación, desde el momento en que inició el paquete hasta su aplicación total (fig. 12-8). Las tasas en las unidades de cuidados intensivos se mantuvieron estables durante 2 años antes de introducir el paquete y disminuyeron progresivamente al 62 % de la tasa de intervención previa después de la intervención, proporcionando evidencia que esta intervención particular logró sus propósitos.

booksmedicos.org

Figura 12-8 ■ Estudio de series temporales. Efectos de un programa para reducir las infecciones por *Staphylococcus aureus* resistente a meticilina (SARM) en instalaciones de Veterans Affairs. Resultados para unidades de cuidados intensivos. (Redibujado con autorización de Jain R, Kralovic SM, Evans ME, et al. Veterans Affairs initiative to prevent methicillin-resistant *Staphylococcus aureus* infections. N Engl J Med 2011;364:1419-1430.)

La inferencia de los estudios de series temporales puede fortalecerse si es posible descartar otras intervenciones que se producen en el mismo momento, como la que se estudió, que podrían haber causado un cambio en las tasas. El caso también se fortalece cuando los resultados intermedios (p. ej., el aumento del lavado de manos) siguen a la intervención (el paquete de SARM), como ocurrió en el ejemplo sobre el SARM.

En un **estudio de múltiples series temporales,** la causa bajo sospecha se presenta en varios grupos diferentes en distintas ocasiones. Posteriormente, se hacen las mediciones de serie temporal en cada grupo para determinar si el efecto se produjo del mismo modo secuencial en el que se presentó la causa bajo sospecha. Cuando un efecto aparece con regularidad tras la introducción de la supuesta causa en diversas ocasiones y en diferentes lugares, existen pruebas más sólidas para la causa que si este fenómeno sólo se ha observado una vez, como en el ejemplo de series temporales únicas, puesto que es incluso más improbable que el mismo factor (o factores) extraño(s) se

booksmedicos.org

produjera(n) en el mismo momento con respecto a la intervención en muchos lugares y momentos diferentes.

Ejemplo

El tamizaje con citología de Papanicolaou se realiza para prevenir las muertes por cáncer del cuello uterino. Esta práctica se inició antes de que hubiera una fuerte evidencia sobre la efectividad, de hecho antes de los ensayos aleatorizados controlados. La mejor evidencia disponible sobre la efectividad proviene de múltiples estudios de series temporales. En la figura 12-9 se muestra un ejemplo (9). En los países nórdicos se aplicaron programas de tamizaje con citología de Papanicolaou en diferentes momentos y con diferentes frecuencias. Las tasas de mortalidad por cáncer del cuello uterino habían aumentado y comenzaron a disminuir justo en los años previos al inicio del tamizaje, ilustrando la importancia de contar con información sobre las tendencias temporales. Las tasas fueron más o menos similares antes de los programas de tamizaje, y disminuyeron más en la mayoría de los países (Islandia, Finlandia y Suecia) con programas nacionales de mayor cobertura y menos en los países (Dinamarca y Noruega) con la menor cobertura.

booksmedicos.org

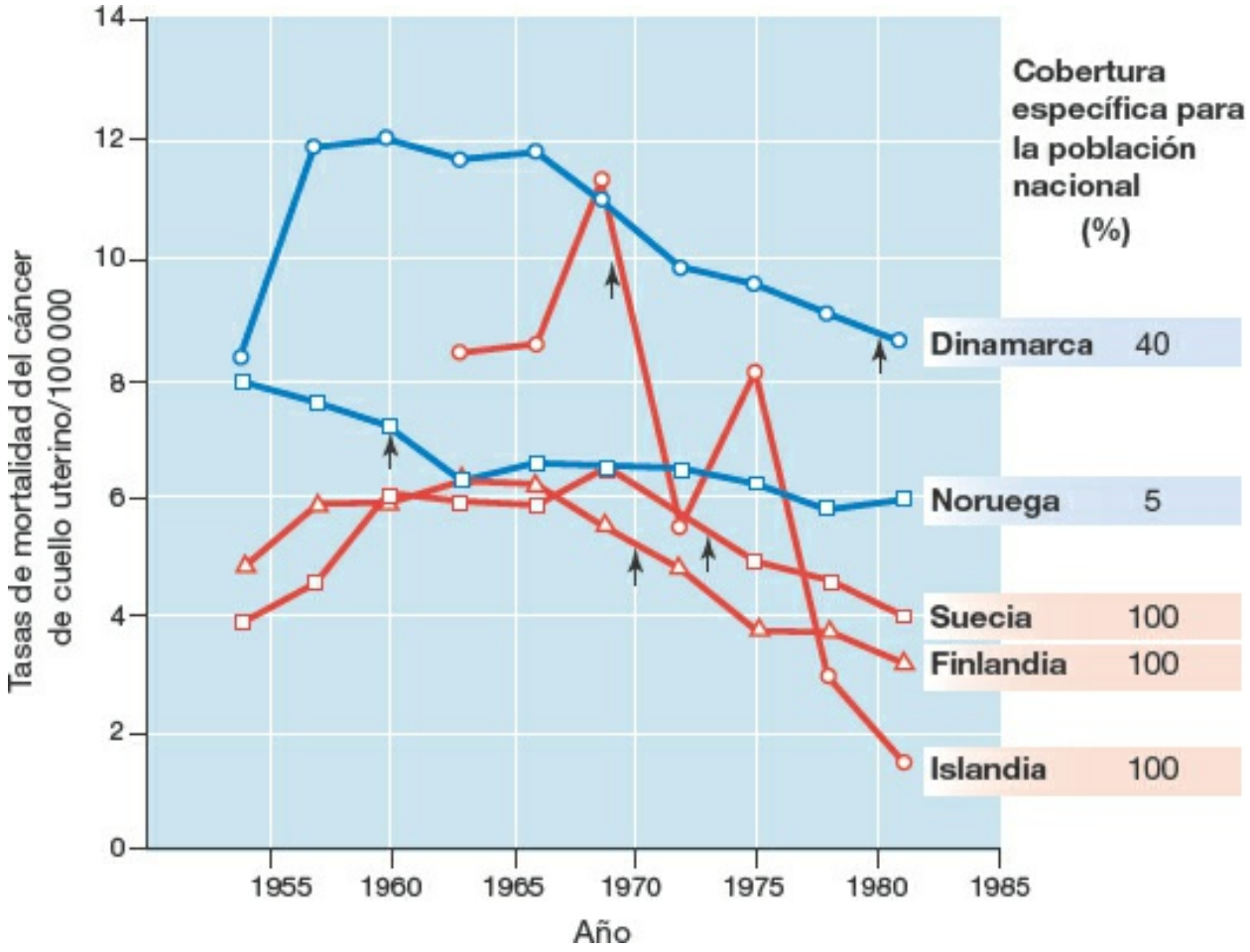


Figura 12-9 ■ Estudio de múltiples series temporales. Cambio en las tasas de mortalidad del cáncer del cuello uterino de acuerdo con el año en el que se aplicaron los programas organizados de tamizaje con citología de Papanicolaou y se cubrió el objetivo. Las *flechas* señalan el año de cobertura que se logró para cada país. (Redibujado con autorización de Lăără E, Day NE, Hakama M. Trends in mortality from cervical cancer in Nordic countries: association with organized screening programmes. Lancet 1987;1[8544]:1247-1249.)

Los estudios de múltiples series temporales como este no suelen ser experimentos planificados. En vez de ello, las intervenciones fueron implementadas en los diferentes países en diferentes momentos y con frecuencias distintas por sus propias razones sociopolíticas. Después, los investigadores tomaron ventaja de este «experimento natural», y de los datos sobre las tasas de mortalidad por cáncer del cuello uterino, para realizar un estudio estructurado.

Los diferentes diseños de los estudios ecológicos tienen una fortaleza científica muy diferente. Relacionar simplemente la exposición añadida con el riesgo añadido entre las diferentes regiones puede ser útil para crear hipótesis que deben ser probadas en estudios más rigurosos. Una sola serie

booksmedicos.org

temporal puede proporcionar una evidencia convincente sobre la causa y el efecto si hay un cambio indiscutible en la tendencia tras la intervención y si se descartan las intervenciones concurrentes. Los estudios de múltiples series temporales pueden proporcionar una fuerte evidencia sobre la causa, posiblemente a la altura de los ensayos aleatorizados, porque es muy improbable que la confusión haya producido los mismos efectos después de la intervención en muchos momentos y lugares diferentes.

MODELADO

Ya se ha hablado de cómo se utilizan los modelos matemáticos para controlar la confusión y desarrollar las reglas de predicción. Otro uso de los modelos es describir la importancia relativa de las diversas causas de enfermedad y su prevención.

Ejemplo

La tasa de mortalidad por cáncer colorrectal en EE.UU. disminuyó un 26 % entre 1975 y el año 2000. ¿Qué provocó este cambio? Los investigadores modelaron las contribuciones respectivas de los cambios en los factores de riesgo, el tamizaje y el tratamiento en la disminución de la tasa (10). Los efectos de la prevención primaria se introdujeron en el modelo como el riesgo relativo tanto para los factores de riesgo (tabaquismo, obesidad, consumo de carnes rojas) como para los factores protectores (aspirina, uso de multivitamínicos y actividad física), junto con la prevalencia de cada uno a lo largo del tiempo. Del mismo modo, se introdujo el tamizaje como las tasas de cada tipo de prueba de tamizaje y sus modificaciones a lo largo del tiempo, y su sensibilidad y especificidad para detectar las lesiones premalignas. Se realizó el modelo de tratamiento, considerando los tipos de pautas de quimioterapia disponibles con el tiempo y su efectividad para reducir la mortalidad. El modelo incluyó los muchos pasos que conducen a los adenomas (precursores del cáncer) hasta el cáncer, el tratamiento y la supervivencia. La figura 12-10 muestra que el modelo se acercó mucho a predecir la disminución real de la mortalidad por cáncer colorrectal. Los cambios en los factores de riesgo representaron el 35 % de la disminución de la mortalidad, el tamizaje otro 53 % de la disminución y el tratamiento sólo un 12 %.

booksmedicos.org

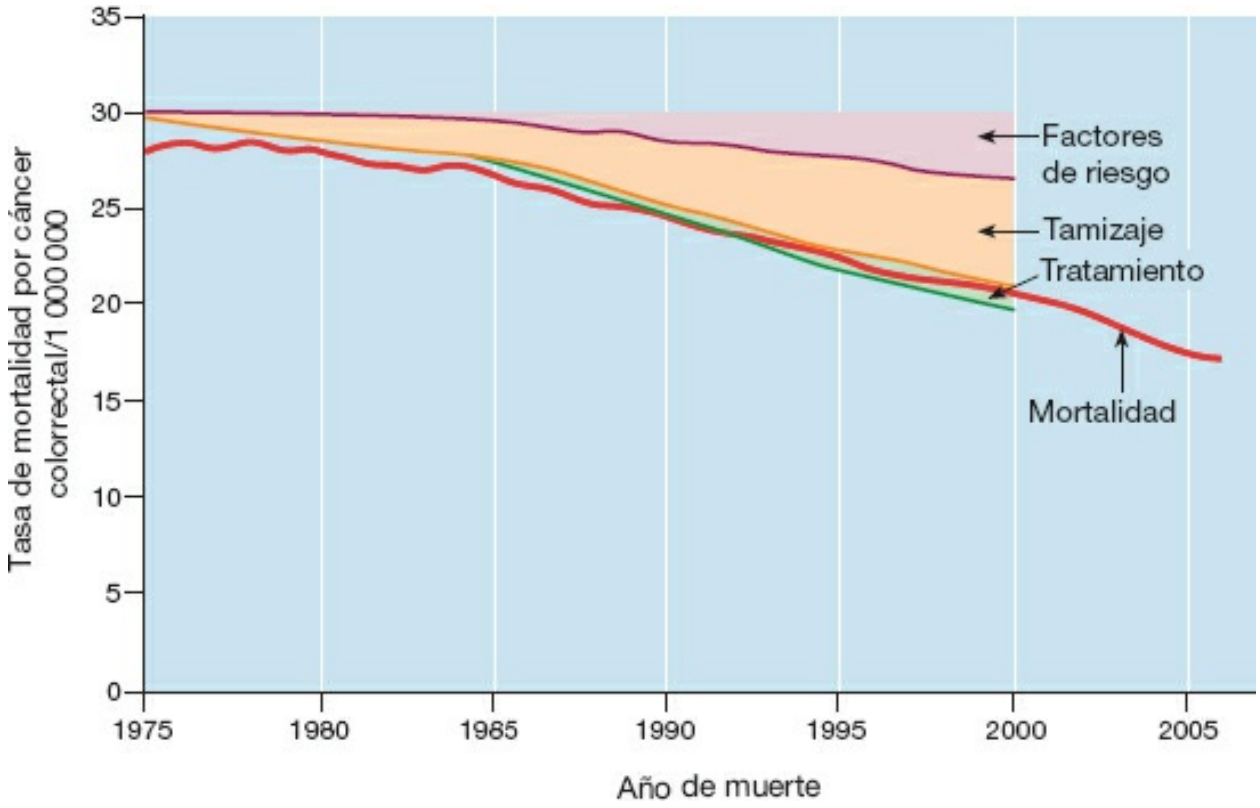


Figura 12-10 ■ Causas de la disminución de las muertes por cáncer colorrectal, 1975-2000. (Redibujado con autorización de Edwards BK, Ward E, Kohler BA, et al. Annual report to the Nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. Cancer 2010;116:544-573.)

Los modelos llevan a un grupo de personas a través de las diversas etapas posibles en la historia natural de la enfermedad. En este caso, fue la población de EE.UU. desde el desarrollo de pólipos hasta su transición al cáncer, la prevención mediante el tamizaje y los efectos del tratamiento en las muertes por cáncer. Los datos sobre las probabilidades de transición de uno a otro estado (p. ej., de los pólipos al cáncer o del cáncer a la curación) se obtuvieron de una investigación previamente publicada.

Otros tipos de modelos se usan para la toma de decisiones cuantitativas, comparando las consecuencias posteriores de las rutas alternativas de acción. El **análisis de decisión** identifica las decisiones que conducen a los mejores resultados en términos humanos, como la supervivencia o el período libre de enfermedad. El **análisis de costo-efectividad** compara el costo con los resultados (p. ej., vida salvada o vidas salvadas ajustadas por la calidad) para los cursos de acción alternativos. En el **análisis de costo-beneficio,** tanto el costo como los beneficios se expresan en términos monetarios. Siempre que el costo se tenga en cuenta en estos modelos, se ajusta según el cambio en el

booksmedicos.org

valor del dinero a lo largo del tiempo, porque se gasta o se ahorra en diferentes momentos durante el curso de una enfermedad. En todos estos modelos, si algunos de los datos ingresados son débiles, se puede utilizar un análisis de sensibilidad para evaluar los efectos de diferentes valores para estos datos.

El modelo puede proporcionar respuestas a las preguntas tan amplias que no estarían disponibles a partir de estudios de investigación individuales, y probablemente nunca lo serán. Se depende cada vez más de ellos para complementar otras formas de investigación cuando se intenta comprender las consecuencias de las decisiones clínicas.

EVALUANDO LA EVIDENCIA

Cuando se determina la causa, se debe considerar la evidencia de todos los estudios disponibles. Tras examinar el diseño de la investigación y la calidad de los estudios, y si sus resultados están a favor y en contra de la causa, se puede fortalecer o debilitar la relación de causalidad.

booksmedicos.org

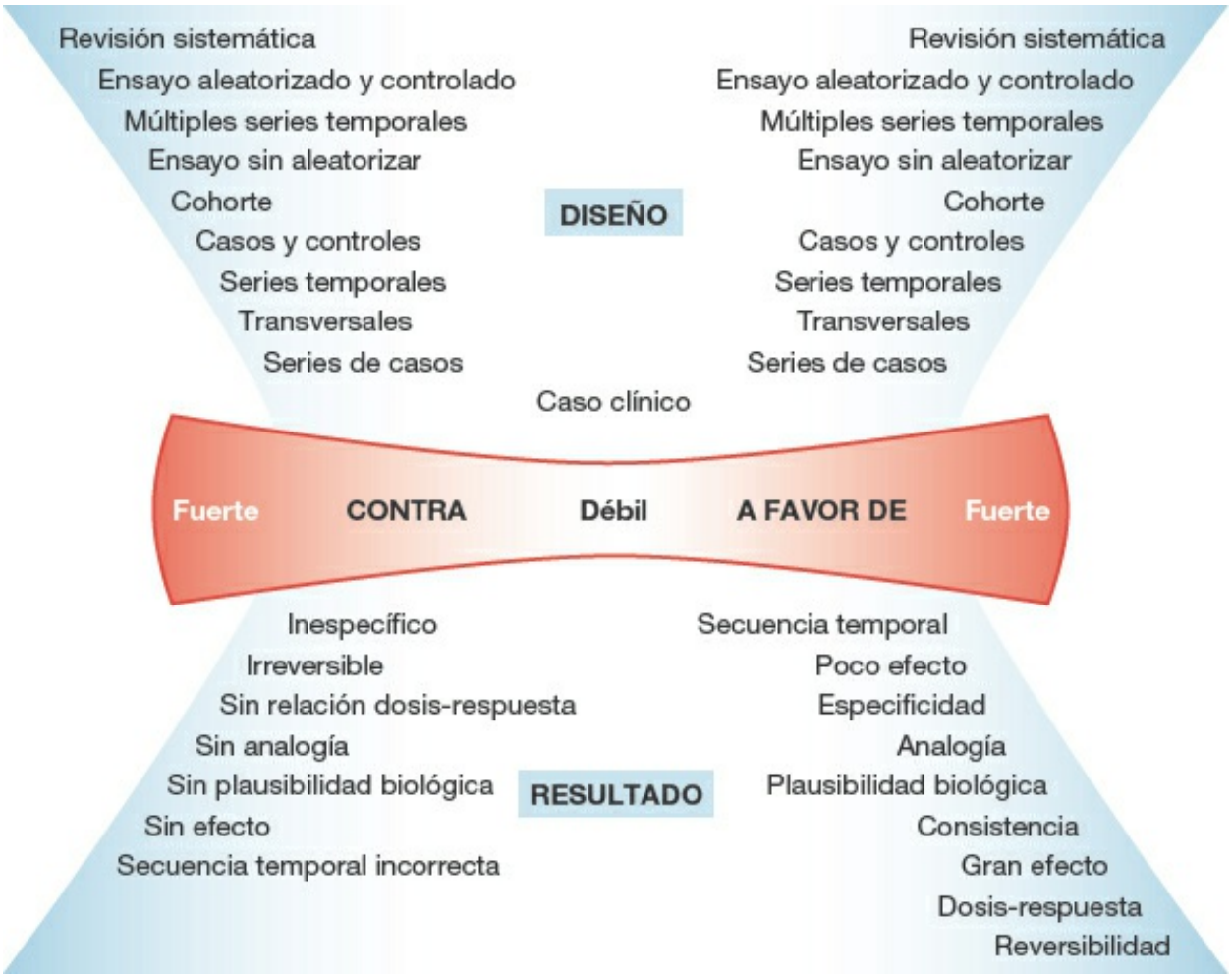


Figura 12-11 ■ Solidez relativa de la evidencia a favor y en contra de un efecto causal. Obsérvese que, con los diseños de estudios, la fuerza de la evidencia a favor de una relación causal es un reflejo de los que están en *contra*. Con los resultados, la evidencia a favor de un efecto causal no refleja la evidencia en *contra* de un efecto.

En la figura 12-11 se resumen los diferentes tipos de pruebas a favor y en contra de la causa, en función del diseño de la investigación y de los resultados que refuerzan o debilitan la evidencia para la causa. Así, un estudio de cohortes minucioso que muestre una asociación sólida y una relación dosis-respuesta reversible constituye una prueba sólida de la causa; en cambio, mientras que un estudio transversal en el que no se ha detectado ningún efecto constituye una prueba poco convincente contra la causa.

La creencia en una relación causa-efecto es un juicio basado en la fortaleza científica y en los resultados de toda la investigación que trata sobre la pregunta. Como un asunto práctico, se debe determinar si el peso de la evidencia es lo suficientemente convincente como para que nos comportemos

booksmedicos.org

como si algo fuese una causa, no si se ha demostrado más allá de toda duda razonable.

Preguntas de revisión

Lea los siguientes enunciados y seleccione la mejor respuesta.

12-1. Uno de sus pacientes leyó que los teléfonos móviles causan cáncer cerebral y quiere saber su opinión. Usted descubre que la incidencia de tumores cerebrales malignos está aumentando en EE.UU. Los resultados de varios estudios observacionales sobre el uso del teléfono móvil y cáncer cerebral son contradictorios. Un ensayo aleatorizado controlado puede resolver la pregunta. ¿Cuál es la razón principal por la que sería poco probable realizar un ensayo aleatorizado controlado (ECA) para esta pregunta?

A. Sería demasiado costoso.
B. Las personas no estarían de acuerdo con la asignación aleatoria al uso del teléfono móvil.
C. Comportaría demasiado tiempo.
D. Incluso si se realizara correctamente, un ensayo aleatorizado no podría responder la pregunta.

12-2. ¿Cuál de las siguientes afirmaciones sería menos compatible con la creencia de que el uso de teléfono móvil es una causa de cáncer cerebral?

A. Una relación dosis-respuesta.
B. Un gran tamaño del efecto (como el riesgo relativo).
C. Un análisis separado de pacientes con tumores del lado derecho o izquierdo de acuerdo con el lado con el que suelen escuchar con su teléfono móvil.
D. Una explicación biológica que aclare por qué los teléfonos móviles podrían causar cáncer.
E. El uso de teléfonos móviles se asocia con muchos tipos diferentes de cáncer.

booksmedicos.org

12-3. ¿Cuál de las siguientes es la descripción más exacta sobre una relación causal?

A. La tuberculosis tiene una causa única, el bacilo de la tuberculosis.
B. La mayoría de las enfermedades genéticas están causadas sólo por un gen anormal.
C. La enfermedad cardíaca coronaria tiene múltiples causas que interactúan entre sí.
D. El tratamiento efectivo ha sido la causa principal de la disminución de las tasas de tuberculosis.

12-4. ¿Cuál de los siguientes *no* es uno de los criterios de causalidad de Bradford Hill?

A. Dosis-respuesta.
B. Significancia estadística.
C. Reversibilidad.
D. Plausibilidad biológica.
E. Analogía.

12-5. Un estudio sobre el tamizaje del cáncer colorrectal que compara la prueba de sangre oculta en heces con la falta de tamizaje encuentra que este cuesta 20 000 dólares por año de vida salvada. ¿A qué tipo de estudio representa este ejemplo?

A. Análisis de costo-beneficio.
B. Análisis de decisión.
C. Análisis de costo-efectividad.
D. Regla de decisión clínica.

12-6. ¿Cuál de las siguientes afirmaciones es verdadera sobre los estudios de riesgo añadido?

A. Los estudios de riesgo añadido no son susceptibles a la confusión.
B. Los estudios de series temporales múltiples no aportan una evidencia fuerte sobre la causa y el efecto.
C. Los estudios que relacionan la exposición promedio con el riesgo promedio pueden proporcionar una fuerte prueba sobre la hipótesis causal.

booksmedicos.org

D. Una falacia ecológica puede afectar a su validez.

12-7. ¿Cuál de las siguientes opciones podría debilitar la creencia de que el «paquete de SARM» (fig. 12-6) causó la disminución observada de las infecciones de SARM en los hospitales VA de cuidados intensivos?

A. El lavado de manos, como parte del paquete de SARM, aumentó después de introducir el programa.

B. Durante la aplicación del programa estuvieron disponibles mejores antibióticos.

C. La tasa de infecciones por SARM se mantuvo estable durante los años previos a la introducción del programa.

D. La disminución en las tasas comenzó justo antes de introducir el programa.

12-8. Los estudios aleatorizados controlados son los diseños de investigación más fuertes para establecer la causa y el efecto. ¿Cuál de las siguientes es la principal limitación de los ensayos para este propósito?

A. Los ensayos aleatorizados no pueden controlar los factores de confusión no medidos.

B. Los ensayos clínicos pueden no ser éticos o factibles para algunas preguntas.

C. El tipo de diseño del estudio es sólo una medida cruda de la fortaleza científica.

D. Los resultados de los ensayos pobremente diseñados y realizados no son mejores que los estudios observacionales.

12-9. ¿Cuál de las siguientes es la evidencia más *fuerte* para una relación causa-efecto??

A. Estudios observacionales que han controlado los sesgos y minimizado el papel del azar.

B. Un mecanismo biológico puede explicar la relación.

C. La supuesta causa precede claramente al efecto.

D. Hay una relación de dosis-respuesta.

E. La evidencia en conjunto es consistente con los criterios de

booksmedicos.org

Bradford Hill.

12-10. Usted descubre que se han realizado estudios de casos y controles para determinar si el uso de teléfono móvil se asocia con el desarrollo de cáncer cerebral. En un estudio se preguntó a los pacientes con cáncer cerebral y a sus controles pareados sin cáncer cerebral sobre el uso de teléfono móvil. El riesgo relativo estimado por al menos 100 h de uso en comparación con la falta de uso fue de 1,0 para todos los tipos de cáncer cerebral combinados (0,6-1,5 intervalo de confianza del 95 %). Este hallazgo es consistente con todo lo siguiente, *excepto:*

A. El uso del teléfono móvil aumenta la incidencia de tumores cerebrales en un 50 %.

B. El uso de teléfonos móviles protege contra los tumores cerebrales.

C. Pueden asociarse tipos específicos de cáncer con el uso del teléfono móvil.

D. El estudio tiene una potencia estadística adecuada para responder a la pregunta de investigación.

Encontrará las respuestas en el apéndice A.

BIBLIOGRAFÍA

1. Holmes OW. On the contagiousness of puerperal fever. *Med Classics* 1936;1:207–268. [Originally published, 1843.]
2. Fouchier RA, Kuiken T, Schutten M, et al. Aetiology: Koch's postulates fulfilled for SARS virus. *Nature* 2003;423:240.
3. MacMahon B, Pugh TF. Epidemiology: Principles and Methods. Boston: Little, Brown & Co.; 1970.
4. Burzynski J, Schluger NW. The epidemiology of tuberculosis in the United States. *Sem Respir Crit Care Med* 2008;29:492– 498.
5. Bradford Hill A. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295–300.
6. Mykletun A, Glozier N, Wenzel HG, et al. Reverse causality in the association between whiplash and symptoms of anxiety and depression. *The HUNT Study. Spine* 2011;36:1380–1386.
7. Moertel CC, Fleming TR, Rubin J, et al. A clinical trial of amygdalin (Laetrile) in the treatment of human cancer. *N Engl J Med* 1982;306:201–206.
8. Jain R, Kralovic SM, Evans ME, et al. Veterans Affairs initiative to prevent methicillin-resistant *Staphylococcus aureus* infections. *N Engl J Med* 2011;364:1419–1430.
9. Lăără E, Day NE, Hakama M. Trends in mortality from cervical cancer in the Nordic countries: association with organized screening programmes. *Lancet* 1987;1(8544):1247–1249.
10. Edwards BK, Ward E, Kohler BA, et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and

booksmedicos.org

treatment) to reduce future rates. *Cancer* 2010; 116:544–573.

booksmedicos.org

Capítulo 13

Resumen de la evidencia

Cuando la comunidad de investigación sintetiza a fondo la evidencia existente, es seguro que cambiará una proporción sustancial de las ideas actuales sobre los efectos de la atención médica. Las formas de atención que actualmente se creían ineficaces demostrarán su eficacia; las formas de atención que se creían útiles serán expuestas como inútiles o perjudiciales, y se pondrá de manifiesto la justificación de la incertidumbre acerca de los efectos de muchas otras formas de atención médica.

—Ian Chalmers y Brian Haynes
1994

PALABRAS CLAVE

Revisiones narrativas
Revisiones sistemáticas
PICO
Sesgo de publicación
Diagrama de embudo
Diagrama de Forest o de bosque
Metaanálisis
Heterogeneidad
Metaanálisis individual
Modelo de efectos fijos
Modelo de efectos aleatorios
Metarregresión
Metaanálisis acumulativo

Las decisiones médicas se basan en el peso de las pruebas en relación

booksmedicos.org

con un tema. Algunas veces, los resultados de grandes estudios son tan convincentes que eclipsan a todos los demás sobre el mismo tema. Sin embargo, muy a menudo los médicos dependen de la acumulación de pruebas de muchos estudios menos definitivos. Cuando se consideran estos estudios individuales, los médicos necesitan establecer el contexto de esa prueba, preguntándose: «¿Ha habido otros estudios sobre el mismo tema, qué han demostrado y sus resultados han establecido un patrón cuando las fortalezas científicas del estudio y la precisión estadística se tienen en cuenta?». Las revisiones están destinadas a responder este tipo de preguntas.

Las revisiones se encuentran a disposición de los usuarios de muchas formas diferentes. Pueden ser artículos en revistas, capítulos en libros de texto, resúmenes preparados por la Cochrane Collaboration o monografías publicadas por organizaciones profesionales o gubernamentales. Si los autores de los artículos de investigación individual están haciendo su trabajo correctamente, proporcionarán información sobre los resultados previos de los estudios en los apartados de introducción o discusión del artículo.

Sin embargo, aunque se cuente con una revisión, el asunto importante es si se ha realizado bien. Existen muchas maneras diferentes, cada una con sus fortalezas y debilidades. En este capítulo describiremos brevemente las revisiones tradicionales y luego abordaremos con mayor detalle un enfoque científico más poderoso y explícito denominado «revisiones sistemáticas».

REVISIONES TRADICIONALES

En las revisiones tradicionales, denominadas **revisiones narrativas,** el especialista con experiencia en la materia resume las pruebas y hace recomendaciones. Una ventaja de estas revisiones es que pueden tratar temas muy diversos, como el «tratamiento del paciente diabético», y abordan una gran cantidad de temas, como criterios diagnósticos, control de glucemia y observación, modificación del factor del riesgo cardiovascular y complicaciones micro y macrovasculares. También se puede incluir el cumplimiento y el costo-efectividad. Los médicos necesitan una orientación respecto a este tipo de cuestiones y los especialistas se encuentran en una buena posición para proporcionarla. Los autores generalmente tienen experiencia con la enfermedad, conocen la

booksmedicos.org

base de las pruebas pertinentes y han aplicado su conocimiento en la atención de los pacientes.

Una desventaja de las revisiones narrativas es que la evidencia y las recomendaciones contenidas en ellas pueden verse influidas por juicios de valor que no son explícitos. La falta de estructura de las revisiones tradicionales puede ocultar amenazas importantes para la validez. Se puede citar la investigación original sin explicitar cómo se encontraron los artículos y, por consiguiente, que sean citados de manera selectiva para apoyar un punto de vista. Se suele incluir la experiencia personal y el saber convencional, y puede resultar difícil distinguir las evidencias sólidas. La validez de la investigación original puede no ser cuidadosamente criticada y, sin embargo, ser mencionada por indicadores de calidad, como el prestigio de la revista, la eminencia del autor, la reciente publicación del estudio, el número de artículos a favor y en contra de una conclusión determinada, y quizá el diseño de la investigación en general (p. ej., un ensayo aleatorizado), sin tener en cuenta si los estudios en realidad se diseñaron y se realizaron bien. Además, puede que no haya una base explícita para determinar por qué un resultado de la investigación se evalúa y otro no.

Por supuesto, una revisión tradicional puede *no* tener estas limitaciones, especialmente si ha sido revisada por otros expertos con experiencia complementaria y si el autor ha incluido la evidencia de revisiones más estructuradas. Sin embargo, la preocupación sobre las limitaciones de las revisiones tradicionales, especialmente la falta de estructura y transparencia, ha promovido un nuevo enfoque.

REVISIONES SISTEMÁTICAS

Las **revisiones sistemáticas** son revisiones rigurosas de la evidencia en relación con temas clínicos específicos. Son «sistemáticas» porque resumen la investigación original siguiendo un plan con base científica establecido por adelantado y se ha explicitado en cada fase. Como consecuencia, los lectores pueden observar la validez de las pruebas para cualquier conclusión a la que se llegue y, en principio, comprobar la autenticidad por sí mismos. En ocasiones, se pueden combinar estudios para obtener una estimación más precisa de la magnitud del efecto que la obtenida en estudios individuales.

booksmedicos.org

Las revisiones sistemáticas son especialmente útiles para abordar exclusivamente un tema, como si los inhibidores de la enzima conversora de la angiotensina reducen el índice de mortalidad de los pacientes con insuficiencia cardíaca congestiva o si los adhesivos cutáneos son mejores que las suturas para los desgarros superficiales. Para que una revisión sistemática sea útil, debe disponerse de estudios importantes sobre el tema. Los estudios sobre el tema no deben ser tan escasos como para que se puedan enjuiciar directamente los estudios concretos y prescindir de la revisión. Los resultados del estudio deben discrepar o al menos dejar la pregunta abierta; si todos los estudios están de acuerdo entre sí, no hay nada que adaptar en la revisión. Las revisiones sistemáticas también son útiles cuando hay motivos para creer que la política, la pasión intelectual y los intereses personales intervienen en la forma de interpretar los resultados de investigación.

Tabla 13-1

Elementos de una revisión sistemática

1. Definir una pregunta específica
2. Encontrar todos los estudios relevantes (publicados o no)
3. Seleccionar los estudios más fuertes
4. Describir la fortaleza científica de los estudios seleccionados
5. Determinar si la calidad se asocia con los resultados
6. Resumir los estudios en diagramas (diagramas de bosque) y tablas
7. Determinar si está justificado combinar los estudios (metaanálisis)
8. De ser así, calcular el tamaño del efecto global y el intervalo de confianza
9. Identificar las causas de heterogeneidad, si está presente

Las revisiones sistemáticas pueden proporcionar una respuesta creíble sobre temas dirigidos (pero no con un objetivo amplio) y ofrecen una serie de posibilidades para realizar mejor las revisiones tradicionales. Estas

booksmedicos.org

complementan, pero no reemplazan, a las revisiones tradicionales. Las revisiones sistemáticas a menudo se usan para resumir los ensayos aleatorizados y controlados; por lo tanto, basaremos nuestros comentarios en los ensayos. Sin embargo, se utilizan los mismos métodos para resumir estudios de riesgo y estudios sobre la realización de una prueba diagnóstica.

Los elementos de una revisión sistemática se resumen en la tabla 13-1, y se abordarán, uno a uno, en este capítulo.

Definir una pregunta específica

Las revisiones sistemáticas tratan sobre preguntas específicas. Para efectos de las intervenciones se han definido los elementos de especificidad bajo el acrónimo **PICO** (1):

P = Pacientes
I = Intervención
C = Comparación
O = Resultados (O del inglés *outcome*)

A estos, algunos han añadido T para el tiempo (p. ej., el tiempo de seguimiento en un estudio de cohorte o ensayo aleatorizado) y S para el diseño del estudio (S del inglés *study design*) para formar PICOTS. Los elementos de las preguntas enfocadas en estudios de diferente tipo (p. ej., estudios sobre la precisión de una prueba diagnóstica o estudios observacionales sobre el riesgo o los factores pronósticos) no están tan bien definidos, pero incluyen muchas de las mismas características.

Encontrar todos los estudios relevantes

El primer paso de una revisión sistemática consiste en reunir todos los estudios que abordan el tema. La revisión debe incluir una muestra completa sobre los mejores estudios sobre el tema, no sólo una muestra sesgada de estudios que le han interesado. Los médicos que revisan temas de una manera más informal (p. ej., en conferencias de colegas, informes y asociaciones de revistas) se enfrentan a un desafío similar y deben utilizar métodos parecidos aunque el proceso raras veces sea tan exhaustivo.

¿Cómo se puede estar razonablemente seguro de que el revisor que él o ella ha reunido los mejores estudios, teniendo en cuenta que hay muchas

booksmedicos.org

publicaciones médicas y están muy dispersas? Ningún método de búsqueda es suficiente para llevar a cabo esta tarea, de modo que se utilizan múltiples enfoques complementarios (tabla 13-2).

La mayoría de las revisiones comienzan con la búsqueda en las bases de datos electrónicas sobre la investigación publicada, entre ellas están MEDLINE (la base de datos electrónica de artículos publicados de la *National Library of Medicine* de EE.UU.), EMBASE y la base de datos de revisiones sistemáticas de Cochrane. Se pueden identificar muchas otras con la ayuda de un bibliotecario. En algunas, como MEDLINE, se puede buscar tanto el área de contenido (como el tratamiento de la fibrilación auricular) como un marcador de calidad (p. ej., un ensayo aleatorizado controlado). Sin embargo, incluso en las mejores manos, la sensibilidad de las búsquedas en MEDLINE (incluso para los artículos que están en MEDLINE) está lejos de ser perfecta. También los contenidos de varias bases de datos tienden a complementarse entre sí. Por lo tanto, la búsqueda en bases de datos es útil pero no suficiente.

Otras formas de encontrar los artículos adecuados compensan lo que la búsqueda en las bases de datos podría omitir. Las revisiones recientes y los libros de texto (particularmente los libros de texto electrónicos que se actualizan continuamente) son otra fuente. Además, los especialistas del tema en cuestión (p. ej., cardiopatía reumática o infección por *Salmonella*) pueden recomendar estudios que los otros enfoques no revelaron. Las referencias que se mencionan en los artículos encontrados son otra posibilidad. Cada vez hay más cantidad de registros de estudios clínicos e investigación financiada que se pueden utilizar para encontrar resultados que no se han publicado.

Tabla 13-2

Procedimientos para encontrar todos los estudios que tratan de un tema en concreto

- Búsqueda en bases de datos electrónicas como MEDLINE, EMBASE y la base de datos de revisiones sistemáticas de Cochrane
- Leer revisiones y libros de texto recientes
- Consultar a especialistas del tema en cuestión
- Considerar los artículos que se mencionan en los trabajos que ya otros

booksmedicos.org

han encontrado
- Examinar los registros de ensayos clínicos y la investigación financiada (para identificar estudios no publicados)

El objetivo de consultar estas fuentes es evitar que se omita cualquier artículo importante, incluso a costa de la ineficacia. En términos de pruebas diagnósticas, el revisor utiliza muchas pruebas paralelas para aumentar la sensibilidad de la búsqueda, incluso a costa de muchos resultados falsos positivos (es decir, no deseados o citas redundantes), que deben eliminarse mediante la consulta de los propios estudios.

Además de practicar la agilidad necesaria para encontrar artículos, los autores de revisiones sistemáticas deben haber descrito de forma explícita la estrategia de búsqueda en su revisión, incluidos los términos de búsqueda. De este modo, los lectores pueden ver en qué medida el revisor consideró todos los estudios relevantes que estaban disponibles en aquel momento.

Limitar las revisiones a estudios científicamente sólidos y clínicamente relevantes

Para estar incluidos en una revisión sistemática, los estudios deben cumplir unos criterios de validez científica. Se supone que sólo deben contar los estudios relativamente importantes. ¿Cómo se establecen los criterios? Varios grupos de expertos han propuesto criterios para la validez científica apropiada, y sus ventajas y limitaciones se discuten más adelante en este capítulo.

Normalmente, de entre un gran número de posibles artículos sobre un tema, sólo se selecciona una pequeña proporción. Muchos artículos describen la biología de la enfermedad y no están preparados para la aplicación clínica. Otros comunican opiniones o resúmenes de pruebas existentes, no de investigaciones clínicas originales. Muchos estudios no son científicamente sólidos y la información que contienen ocultan otros estudios mejores. Relativamente pocos reportan evidencia que tenga relación directa con la pregunta clínica y que sea tanto científicamente sólida como clínicamente relevante. La tabla 13-3 muestra cómo se seleccionaron los artículos de una revisión sistemática sobre las estatinas para la prevención de infecciones; sólo 11 de 632 publicaciones

booksmedicos.org

identificadas fueron incluidas en la revisión.

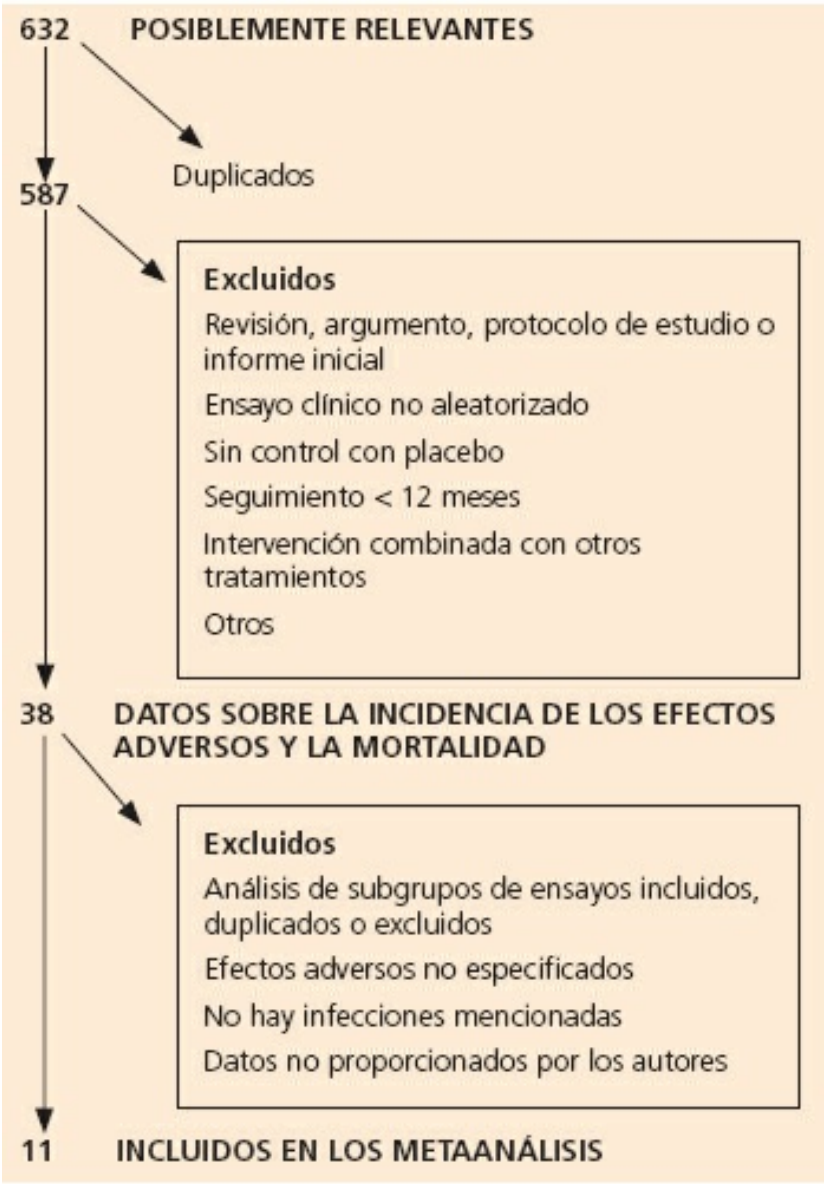
¿Son los estudios publicados una muestra sesgada de toda la investigación concluida?

Los artículos a los que se alude en las revisiones sistemáticas deben contener todos los estudios científicamente importantes del tema en cuestión, sin importar si han sido publicados o no. El **sesgo de publicación** es la tendencia de que los estudios publicados sean sistemáticamente diferentes de todos los estudios completos sobre un tema. En general, es más probable que los estudios publicados sean «positivos» (tengan un efecto) por varias razones, relacionadas con la preferencia general a resultados positivos. Es menos probable que los investigadores terminen estudios que parecen propensos a tener resultados negativos y que se presenten estudios negativos en las revistas. Es menos probable que los colegas revisores de las revistas consideren que un resultado negativo es una noticia interesante y los editores son menos propensos a publicarlos.

booksmedicos.org

Tabla 13-3

Las revisiones sistemáticas sólo integran una pequeña proporción de los artículos sobre un tema. Artículos considerados e incluidos en una revisión sistemática sobre las estatinas y la prevención de infecciones



De Van den Hoek HL, Bos WJ, de Boer A, et al. Statins and the prevention of infections: systematic review and meta-analysis of data from large randomized placebo controlled trials. BJM 2011;343:d7281.

booksmedicos.org

Otra influencia selectiva puede favorecer a los estudios positivos. Los autores pueden informar sobre resultados que no se identificaron o se asumieron de antemano, antes de recoger los datos, y que fueron seleccionados por su énfasis en las publicaciones después de contar con los resultados.

Con el fin de sortear estos problemas, algunos autores de revisiones sistemáticas realizan un esfuerzo para conseguir estudios que no se han publicado, incluidos aquellos que recibieron subvenciones pero que no se terminaron. Para llevar a cabo esta tarea cuentan con la ayuda de los registros públicos de todos los estudios que se han comenzado.

Los **diagramas de embudo** son un método gráfico para detectar sesgos en la selección de estudios para las revisiones sistemáticas. Para cada estudio se dibuja el tamaño del efecto contra alguna medida del tamaño o precisión del estudio, como el tamaño de la muestra, el número de efectos de resultado o el intervalo de confianza (IC) (fig. 13-1).

Si no existe sesgo de publicación (fig. 13-1 A), es más probable que los estudios grandes (representados en la parte superior de la figura) sean publicados sin importar los hallazgos y produzcan estimaciones de la efectividad comparativa que se agrupan estrechamente alrededor del verdadero tamaño del efecto. Es más probable que el tamaño del efecto reportado varíe en los estudios pequeños (representados en la parte inferior de la figura) debido a la imprecisión estadística y, por lo tanto, que puedan extenderse en la parte inferior de la figura, rodeando al verdadero tamaño del efecto. En ausencia del sesgo de publicación, estarían a menudo tanto en la parte inferior derecha como en la parte inferior izquierda de la figura. El resultado, si no hay sesgos, es una distribución picuda simétrica (un embudo invertido).

El sesgo de publicación, especialmente la tendencia de publicar los estudios pequeños sólo si son positivos, nos muestra asimetría en el diagrama de embudo (fig. 13-1 B). Hay un número mucho menor de estudios que favorecen al grupo control, lo que se evidencia por la escasez de estudio en la esquina inferior derecha de la figura.

Otros factores que no están directamente relacionados con el desagrado por estudios negativos pueden causar un sesgo de publicación. Las investigaciones financiadas por agencias con un interés económico en los resultados también pueden causar distorsiones en el registro científico. Es habitual que los resultados de los estudios financiados por la industria

booksmedicos.org

(normalmente por empresas farmacéuticas y de aparatos) favorezcan más al producto de la compañía que los patrocina que a aquellos con otras fuentes de subvención. Un posible motivo es que las empresas patrocinadoras suelen requerir, como condición para recibir los fondos, que sean ellas las que aprueben los artículos antes de que se envíen a las revistas. Estas empresas han bloqueado la publicación de las investigaciones que han subvencionado, ya que en su opinión no han logrado el resultado «correcto».

booksmedicos.org

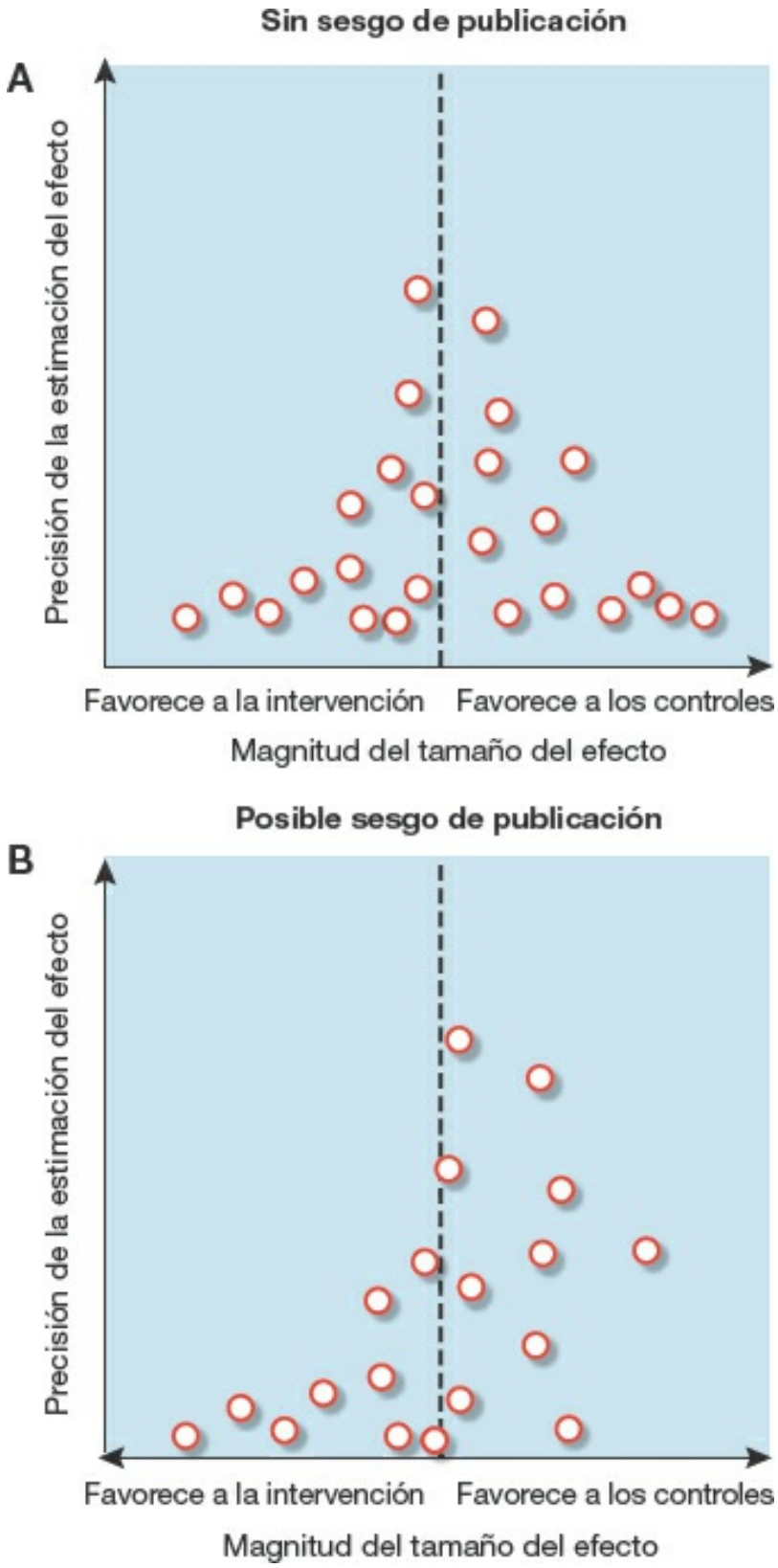


Figura 13-1 ■ Diagramas de embudo para detectar sesgos de publicación. Cada ensayo está indicado por un círculo. **A.** Los ensayos son simétricos en la forma de un embudo invertido, lo que

booksmedicos.org

sugiere que no hay sesgos en la publicación. **B.** No hay ensayos en la esquina inferior derecha, lo que sugiere que los ensayos pequeños que no favorecen la intervención no se publicaron. (Redibujado con autorización de Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence = publication bias. J Clin Epidemiol 2011;64:1277-1282.)

¿Cuál es la calidad de los mejores estudios?

Los médicos deben conocer la calidad de los mejores estudios que existen sobre un tema para saber la seriedad con la que se han de tomar las conclusiones de una revisión sistemática. ¿Son los estudios tan relevantes que sería irresponsable descartarlos? ¿O son tan insignificantes que no es conveniente seguir el ejemplo?

Muchos estudios han mostrado que los elementos individuales de calidad comentados a lo largo de este libro, como el encubrimiento de la asignación del tratamiento, el cegamiento, el seguimiento y el tamaño de la muestra, se relacionan sistemáticamente con los resultados del estudio. La calidad de la evidencia identificada a través de una revisión sistemática puede resumirse en una tabla que muestra el grado en que los marcadores de calidad están presentes en los estudios incluidos en la revisión.

Ejemplo

Una revisión sistemática resumió los informes publicados sobre la efectividad del suplemento condroitina en la intensidad del dolor en los pacientes con osteoartritis de cadera o rodilla (2). Los investigadores encontraron 20 ensayos publicados que compararon la glucosamina con el placebo. La figura 13-2 muestra la proporción de ensayos que cumplen con varios criterios de calidad. (Ignore, por el momento, los tamaños del efecto, que se muestran a la derecha.) Una proporción sustancial de ensayos no cumplieron con los estándares básicos de calidad científica, como el encubrimiento de la asignación, el análisis por intención a tratar y el tamaño de la muestra, por lo que la base de evidencia para este tema no fue particularmente fuerte.

Las medidas individuales de calidad también pueden combinarse en medidas de resumen. Una escala simple de uso frecuente para los estudios sobre la efectividad del tratamiento, la Escala de Jadad, considera si el estudio se describió como aleatorizado y con cegamiento doble, y si se describieron las pérdidas y los abandonos (3). Sin embargo, no existe una relación clara entre las puntuaciones de resumen para la calidad y los resultados del estudio (4). ¿Por qué podría suceder esto? Los estudios que componen las revisiones sistemáticas ya están altamente seleccionados y,

booksmedicos.org

por lo tanto, la calidad entre uno y otro podría no diferir mucho. Además, las medidas de resumen sobre la calidad normalmente suman puntuaciones por la presencia o ausencia de cada elemento de calidad, y no hay razón para creer que cada uno hace una contribución equitativa a la validez general del estudio. No es difícil imaginar, por ejemplo, que la debilidad de un aspecto del estudio podría ser tan perjudicial como para invalidarlo por completo, a pesar de que todos los demás aspectos de calidad sean ejemplares.

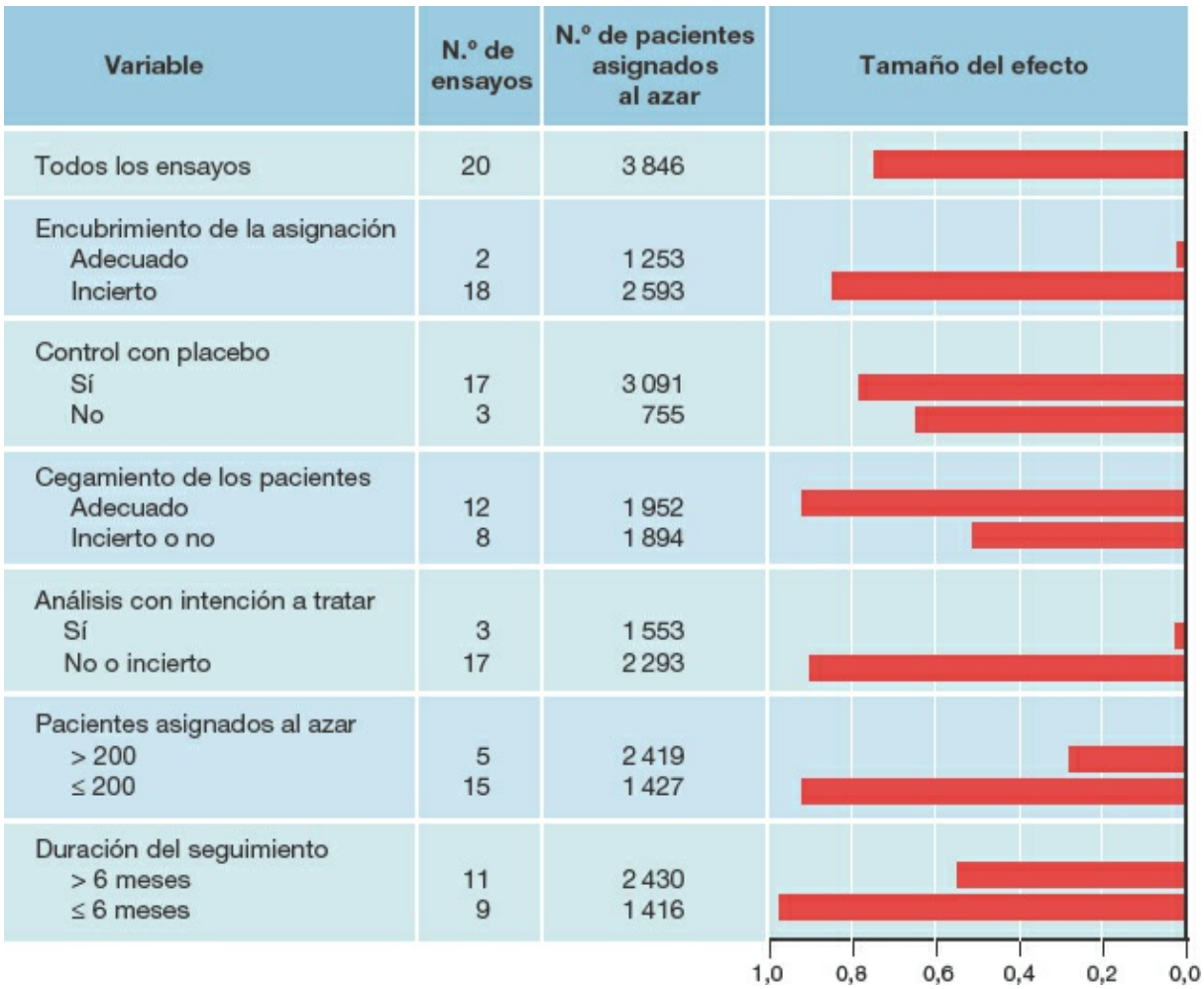


Figura 13-2 ■ Calidad de 20 ensayos en una revisión sistemática sobre la efectividad de la glucosamina para el dolor en pacientes con osteoartritis de rodilla o cadera. (De Reichenbach S, Sterchi R, Scherer M, et al. Meta-analysis: chondroitin for osteoarthritis of the knee and hip. Ann Intern Med 2007;146:580-590.)

booksmedicos.org

Ejemplo

En un ensayo clínico aleatorizado y controlado con placebo, las mujeres con fatiga idiopática pero sin anemia se encontraban menos cansadas tras ingerir un suplemento de hierro (5). El efecto fue pequeño, alrededor de 1 punto en una escala de fatiga de 10 puntos, pero estadísticamente significativo. La mayoría de los aspectos del estudio como la aleatorización, el encubrimiento de la asignación, el tamaño muestral y el seguimiento eran convincentes. Además, los comprimidos de hierro y de placebo tenían la misma apariencia. Sin embargo, el hierro causa efectos secundarios característicos (como heces negras y estreñimiento), y por lo tanto las mujeres podrían haberse dado cuenta de si estaban tomando hierro o placebo. Además, la fatiga es un resultado relativamente «leve» en el que pueden influir las creencias sobre la eficacia de los aportes del hierro. Aunque todos los demás aspectos fueran irreprochables (y la puntuación de la calidad, excelente), este aspecto de calidad (pacientes que conocen el tratamiento que están tomando junto con un resultado fácilmente influenciable por este conocimiento) pudo haber sido la causa de la pequeña diferencia que se observó, más que un efecto real del tratamiento.

Por lo tanto, aunque las listas de cotejo de calidad y las puntuaciones tienen su lugar, no sustituyen a los estudios por separado de una revisión sistemática que analizan de forma crítica, con la vista puesta en cualquier posible imperfección que tengan los estudios y que pueda haber influido en los resultados.

¿La calidad científica está relacionada con los resultados de la investigación?

Los estudios que cumplen con los métodos más estrictos deben acercarse más a la verdad que los más débiles. Por lo tanto, puede ser informativo para los revisores demostrar si existe una relación entre la calidad del estudio y sus conclusiones. Esto se suele hacer examinando la relación entre las medidas de calidad individual y los resultados debido a las limitaciones de las puntuaciones de resumen.

Ejemplo

La figura 13-2 muestra la relación entre varios marcadores de calidad y tamaño del efecto, con intervalos de confianza, para los ensayos de condroitina contra placebo o ninguna intervención para el alivio del dolor en la osteoartritis de rodilla o cadera (2). En general, los tamaños del efecto fueron mayores en los ensayos en los que no hubo una medida de calidad; este efecto

booksmedicos.org

fue estadísticamente significativo para el encubrimiento de la asignación, el análisis por intención a tratar y los estudios más grandes. Un análisis restringido a los tres estudios más grandes con análisis por intención a tratar encontró un tamaño del efecto de -0,03 (IC del 95 % de -0,13-0,07). Se consideró clínicamente relevante un tamaño del efecto mínimo de -0,30. Es decir, no hubo un efecto clínico importante.

Resumir los resultados

Los resultados de una revisión sistemática suelen presentarse como un **diagrama de Forest o de bosque,** en el que se muestran la estimación puntual de la efectividad y el IC de cada uno de los estudios en la revisión. La figura 13-3 ilustra el resumen de los estudios que comparan la quinina con el placebo para el tratamiento de los calambres musculares (6). La medida de la eficacia de este ejemplo es un cambio en el número de calambres musculares en 2 semanas. En otras revisiones sistemáticas podría ser el riesgo relativo, el riesgo atribuible o cualquier otra medida del efecto. Las estimaciones puntuales se representan mediante recuadros con un tamaño proporcional al tamaño del estudio. Una línea vertical marca el punto donde no fueron más efectivos ni la quinina ni el placebo.

booksmedicos.org

Estudio o subgrupo	N.º de calambres	Peso	N.º de calambres (IC del 95 %)
CIBA 1988	–1,68	5,7 %	
Diener 2002	–3	8,7 %	
Fung 1989	–3,69	13,0 %	
Gorlich 1991	–2,69	0,4 %	
Hays 1986	–0,55	15,5 %	
Jansen 1994	–2,7	3,4 %	
Jansen 1997	–6	6,1 %	
Jones 1983	–2	2,4 %	
Lee 1991	–3,55	11,1 %	
Leo Winter 1986	–1,26	14,5 %	
Sidorov 1993	–0,7	8,5 %	
Warburton 1987	–2,18	9,5 %	
Woodfield 2005	–14,2	1,2 %	
Total (IC del 95 %)		**100,0 %**	

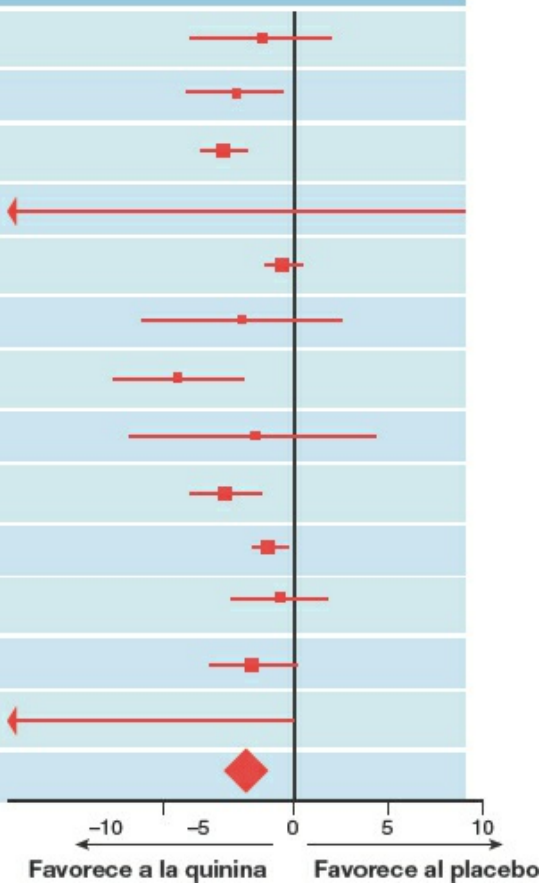


Figura 13-3 ■ Ejemplo de una gráfica de bosque. Resumen de 13 ensayos aleatorizados sobre la efectividad de la quinina contra placebo en el número de calambres en 2 semanas. (Redibujado con autorización de El-Tawil S, Al Musa T, Valli H, et al. Quinine for muscle cramps. Cochrane Database Syst Rev 2010;[12]:CD005044.)

El origen del nombre «diagrama de bosque» es incierto, pero en general se atribuye al nombre del investigador o a la apariencia, ya que parece «un bosque de líneas» (7). Nosotros pensamos que ayuda a los lectores a «ver el bosque y los árboles».

Los diagramas de bosque proporcionan a los lectores una gran cantidad de información que requeriría un gran esfuerzo para encontrarla de otro modo.

1. *Número de estudios*. Las filas muestran el número de estudios que cumplen con los criterios rigurosos de calidad, en este caso, 13.
2. *Qué estudios y cuándo*. La primera columna identifica los nombres y el año de publicación de los estudios que lo forman, para que los lectores

booksmedicos.org

puedan observar los años que tienen los estudios y dónde pueden encontrarlos. (Las referencias a los estudios no se muestran en la figura, pero se incluyen en el artículo.)

3. *Diseño de las magnitudes del efecto*. Las 13 estimaciones puntuales, consideradas en conjunto, muestran los diversos estudios reportados por los tamaños del efecto. En este ejemplo, los 13 estudios favorecieron a la quinina, pero el tamaño del efecto varía.
4. *Precisión de las estimaciones*. Muchos estudios (6 de 13) fueron «negativos» (sus intervalos de confianza no incluían al efecto). Esto podría dar la impresión, contando simplemente el número de estudios «positivos» y «negativos», que el tratamiento no es efectivo o, al menos, su efectividad es cuestionable. El diagrama de bosque da una impresión diferente: todas las estimaciones puntuales favorecen a la quinina, y los estudios negativos tienden a ser imprecisos pero consistentes con la efectividad.
5. *Efectos para los estudios grandes*. Los estudios grandes con precisión estadística (que se identifican tanto por intervalos de confianza estrechos como por casillas grandes que representan las estimaciones puntuales) merecen más peso que los pequeños. En el ejemplo, los intervalos de confianza para los dos estudios más grandes no incluyen «ningún cambio en la tasa de calambres musculares» (aunque uno lo aborda).

De este modo, una sola ilustración transmite de inmediato mucha información básica sobre los mejores estudios de un tema en concreto.

COMBINACIÓN DE ESTUDIOS EN METAANÁLISIS

El **metaanálisis** consiste en combinar («agrupar») los resultados de estudios individuales si son los suficientemente similares para justificar un resumen cuantitativo del tamaño del efecto. Cuando es adecuado, el metaanálisis proporciona estimaciones más concretas de magnitudes de efectos que estén disponibles en cualquiera de los estudios por separado.

¿Son los estudios lo suficientemente similares como para justificar la combinación?

No tiene sentido reunir los resultados de estudios muy distintos: estudios de diferentes tipos de pacientes, intervenciones quirúrgicas, seguimientos y

booksmedicos.org

resultados. El hecho de considerar las «manzanas y naranjas» como si ambas fueran sólo frutas pasa por alto información útil.

Los investigadores usan dos enfoques generales para decidir si es apropiado agrupar los resultados del estudio. Una de ellas es hacer un juicio informado sobre si las preguntas de investigación abordadas por los ensayos son lo suficientemente similares como para constituir estudios de la misma pregunta (o un conjunto de preguntas razonablemente similares).

Ejemplo

¿Los suplementos antioxidantes previenen los carcinomas gastrointestinales? Los investigadores buscaron la respuesta a esta pregunta con una revisión sistemática (8). Los antioxidantes fueron las vitaminas A, C, E y B6, β-caroteno (que tiene actividad de vitamina A) y los minerales zinc y selenio, solos y combinados en diferentes dosis. Los carcinomas fueron de esófago, estómago, colon, recto, páncreas e hígado. Los investigadores calcularon una estimación combinada del riesgo para los 14 ensayos aleatorizados que encontraron. La estimación puntual estuvo cerca de 1 (sin efecto) y el IC fue estrecho, lo que indica que no hay un efecto global. Pero, ¿deberían haber combinado estos estudios? A primera vista, las preguntas del estudio eran muy diferentes entre sí. Aunque las diversas vitaminas y minerales tienen en común que son antioxidantes (al igual que muchos otros componentes de la dieta), tienen diferentes estructuras químicas, funciones en el metabolismo y síndromes a causa de su deficiencia. De forma similar, los carcinomas gastrointestinales tienen diferente genética, factores de riesgo y patogénesis. Por ejemplo, el cáncer gástrico se debe a la interacción de la bacteria *Helicobacter pylori*, mientras que el cáncer hepático es a menudo una complicación de la hepatitis viral. Parece improbable que estos suplementos y estos cánceres tengan suficiente en común para justificar tratarlos simplemente como «suplementos antioxidantes» y «carcinomas gastrointestinales» y encontrar un resultado significativo. Que los investigadores no hayan encontrado ningún efecto puede ser debido a las diferencias entre los estudios y no descarta los efectos de suplementos específicos en carcinomas específicos.

Las personas sensatas pueden discrepar sobre la similitud de los estudios para que se justifique la combinación. Hemos criticado la combinación en este estudio, pero otras personas competentes pensaron que la combinación estaba justificada y que era digno de publicarse en una revista destacada.

Otro enfoque es utilizar una prueba estadística para valorar la **heterogeneidad,** el grado en que los resultados del estudio son diferentes el uno del otro, más allá de lo que podría esperarse por el azar. Aquí, la hipótesis nula es que no existe diferencia entre los resultados del estudio y la prueba estadística se utiliza para ver si hay una diferencia

booksmedicos.org

estadísticamente significativa entre ellos. No conseguir rechazar la hipótesis nula de que no existe diferencia alguna entre los estudios puede ser tranquilizador, pero hay un problema. La mayor parte de los metaanálisis incluyen relativamente pocos estudios y, por lo tanto, las pruebas tienen una potencia estadística limitada. El riesgo de obtener un resultado falso negativo, una conclusión que los estudios no son heterogéneos cuando lo son, es bastante habitual. La potencia también se ve afectada por la cantidad de pacientes y la uniformidad con la que se distribuyen en estos estudios. Si uno de los estudios integrantes es mucho mayor que los otros, contribuye con la mayor parte de la información relacionada con el tema. Puede ser más informativo examinar un estudio grande cuidadosamente y luego compararlo con los otros. En el ejemplo de los antioxidantes y los carcinomas gastrointestinales, la prueba estadística demostró heterogeneidad.

¿Qué son estudios o pacientes combinados?

Hasta ahora se ha comentado cómo se combinan los *estudios*, que es el procedimiento habitual de la realización de los metaanálisis. Sin embargo, un método todavía más convincente consiste en obtener datos sobre cada *paciente* concreto en cada uno de los estudios integrantes, y reunirlos para producir, de hecho, un gran estudio único denominado **metaanálisis individual** (a nivel del paciente). Relativamente pocos metaanálisis se realizan de este modo, debido a las dificultades para conseguir todos estos datos de muchos investigadores diferentes y conciliar la manera de codificar las variables. Sin embargo, cuando se puede hacer un metaanálisis individual, es posible buscar efectos en subgrupos de pacientes clínicamente importantes. Es probable que la cantidad de pacientes de estos subgrupos sea demasiado pequeña en los estudios por separado para que se obtengan estimaciones estables de los efectos, pero puede ser suficientemente grande cuando se combinan los pacientes de diversos estudios.

Ejemplo

El óxido nítrico inhalado es efectivo en los neonatos de término con hipertensión pulmonar e insuficiencia respiratoria hipóxica. Sin embargo, se dudaba si el óxido nítrico era efectivo en

booksmedicos.org

neonatos prematuros. Los investigadores realizaron un metaanálisis individual sobre los ensayos de óxido nítrico en neonatos prematuros (9). Combinaron los datos de 3 298 neonatos prematuros en 12 ensayos y no encontraron un efecto sobre la mortalidad o la enfermedad pulmonar crónica (59 % frente a 61 % a favor del óxido nítrico, riesgo relativo 0,96, IC del 95 % de 0,92-1,01). Debido a que había datos para cada paciente, fue posible buscar la efectividad en subgrupos clínicamente relevantes de neonatos prematuros, así como se podría con un solo ensayo grande. No hubo diferencias en la efectividad en relación con la edad gestacional, el peso al nacer, los nacimientos múltiples, la raza, el uso de esteroides prenatales y otras siete características de los neonatos.

¿Cómo se combinan los resultados?

Cuando se combinan los resultados del estudio, cada estudio individual contribuye al efecto global en relación con su tamaño (en sentido estricto, el inverso de la varianza). Aquellos que contribuyen con grandes cantidades de información tienen un mayor peso que los que hacen pequeñas contribuciones. Esto es evidente en el ejemplo de la quinina (fig. 13-3), donde se informa del peso de cada estudio en la tercera columna, y el total suma un 100 %. Cuatro de los estudios más grandes contribuyen más al efecto global que los otros nueve estudios más pequeños.

Para resumir los estudios de un metaanálisis se utilizaron dos tipos de modelos matemáticos. Estos modelos difieren en lo que se resume y en lo prudentes que son para calcular los intervalos de confianza globales.

Con el **modelo de efectos fijos** (fig. 13-4 A) se supone que cada estudio trata exactamente del mismo tema, de modo que los resultados de los estudios se diferencian sólo por el azar. Este modelo se denomina «efecto fijo» porque se asume que sólo hay un tamaño del efecto subyacente para todos los estudios, aunque los resultados de los estudios por separado difieran entre ellos debido al azar.

El problema principal de este enfoque es que, a primera vista, los estudios raras veces se parecen tanto en cuanto a pacientes, intervenciones, seguimiento y resultados que se puedan considerar simples réplicas unos de otros. ¿Realmente debe considerarse que las vitaminas A, C y E y el selenio son simplemente ejemplos de «suplementos antioxidantes», aun cuando tienen diferente estructura bioquímica y mecanismo de acción? ¿O bien son lo suficientemente diferentes unos de otros que pueden tener efectos distintos? En la medida en que los temas del estudio se diferencian en algo, la amplitud de los intervalos de confianza globales calculados mediante el modelo de efecto fijo tiende a significar un grado mayor de

booksmedicos.org

precisión de lo que sucede en realidad. Además, mediante la combinación de estudios distintos, se pierde información útil que podría haber surgido al contrastarlos. El modelo del efecto fijo se usa cuando los estudios de una revisión sistemática son claramente bastante similares entre sí.

booksmedicos.org

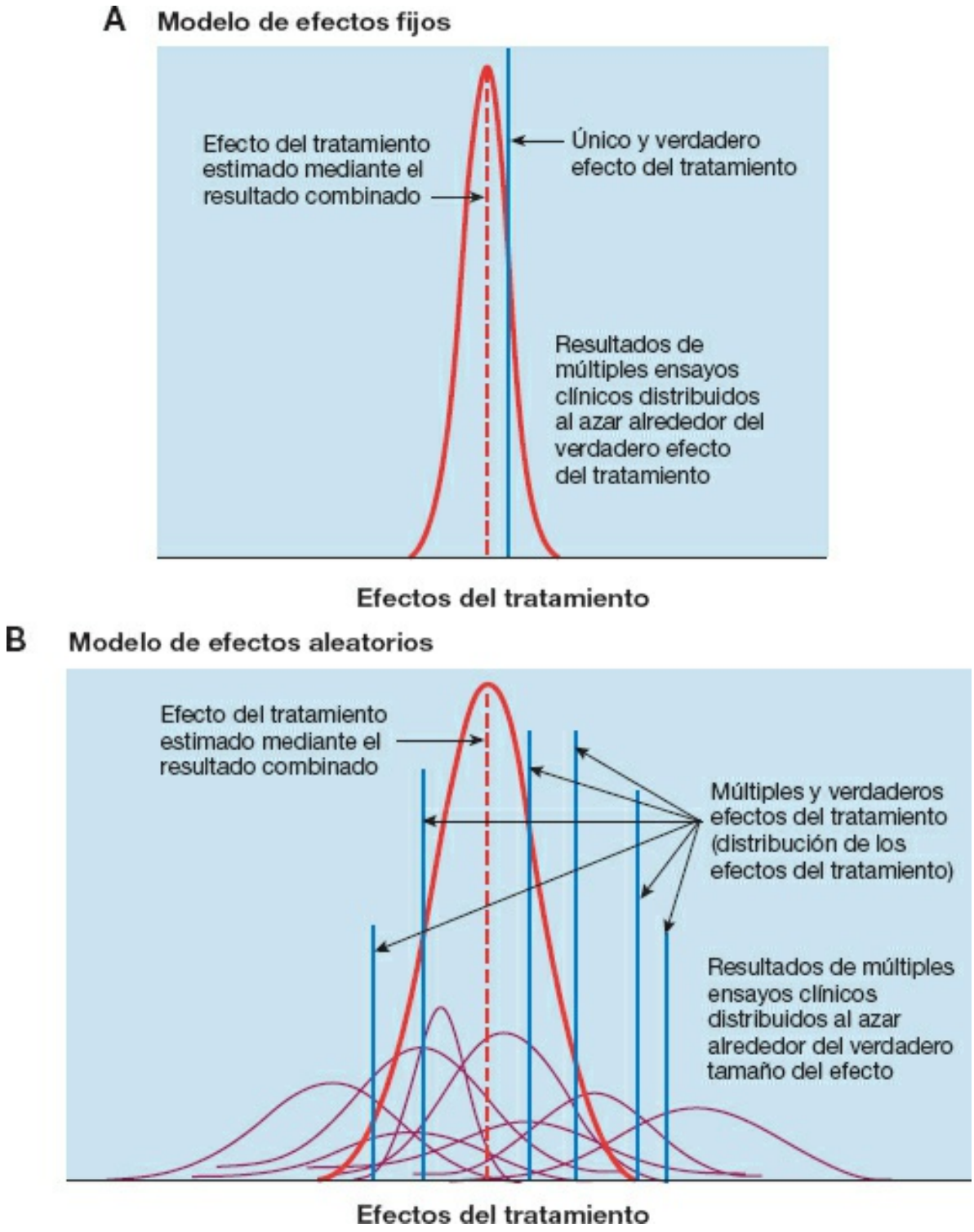


Figura 13-4 ■ Modelos que combinan los estudios en un metaanálisis. A. Modelo de efecto fijo. **B.** Modelo de efectos aleatorios. (Redibujado con autorización de UpTo-Date, Waltham, MA.)

El **modelo de efectos aleatorios** (fig. 13-4 B) supone que los estudios se dirigen a temas algo diferentes y que forman una familia de estudios

booksmedicos.org

estrechamente relacionados sobre un tema similar. Los trabajos se consideran una muestra aleatoria de todos los estudios sobre ese tema. Incluso si el juicio clínico y las pruebas estadísticas sugieren heterogeneidad, puede seguir siendo razonable combinar estudios utilizando un modelo de efectos aleatorios, siempre que los estudios sean lo suficientemente similares entre sí (evidentemente, una valoración válida). Los modelos de efectos aleatorios producen intervalos de confianza más amplios que los modelos de efectos fijos y, por este motivo, se consideran más realistas. Sin embargo, no se sabe con certeza cómo se define la familia de estudios similares ni si los estudios son realmente una muestra aleatoria de todos los trabajos sobre un tema. A pesar de ello, dado que los modelos de efectos aleatorios por lo menos contemplan la heterogeneidad, y es menos probable que sobrevaloren la precisión, son el modelo utilizado con mayor frecuencia en los últimos años.

Cuando se calcula una magnitud global del efecto, se suele presentar como una estimación puntual y un IC en la parte inferior del diagrama de bosque de los estudios componentes como un diamante que representa la estimación puntual resumida y el IC (v. fig. 13-3).

Identificar las causas de heterogeneidad

Los modelos de efectos aleatorios son una forma de tomar en consideración la heterogeneidad cuando se calcula un tamaño del efecto global, aunque también es necesario identificar las características de los pacientes o los tratamientos que son responsables por la variación en los efectos.

La forma más sencilla de identificar las causas de heterogeneidad es hacer un análisis de subgrupos. Esto es posible en los metaanálisis individuales, como se describió en el ejemplo sobre el tratamiento con óxido nítrico en los neonatos prematuros. Sin embargo, si se combinan los ensayos y no los pacientes, se debe confiar en métodos menos directos.

Otro enfoque para comprender las causas de heterogeneidad es hacer un análisis de sensibilidad, como se comentó en el capítulo 7. Se examinan los efectos globales con y sin los ensayos que parecen ser diferentes de los otros, ya sea por motivos clínicos o estadísticos. Por ejemplo, los investigadores pueden observar los efectos globales (o las pruebas estadísticas para la heterogeneidad) después de retirar los ensayos relativamente débiles, o aquellos en los que la dosis del fármaco fue

booksmedicos.org

relativamente pequeña para determinar si la fuerza del estudio o la dosis del fármaco explican los diferentes resultados entre los estudios.

Un enfoque de modelo denominado **metarregresión,** similar al análisis multivariable (comentado en el capítulo 11), se puede usar para explorar las causas de heterogeneidad cuando se combinan los ensayos, no los pacientes. Las variables independientes son las que se informan de forma agregada en cada ensayo individual (p. ej., la edad promedio o la proporción de hombres y mujeres en esos ensayos) y los resultados son el efecto del tratamiento informado para cada uno de los ensayos. El número de observaciones es el número de ensayos en el metaanálisis. Este enfoque está limitado por la disponibilidad de los datos sobre las covariables de interés en los ensayos individuales, la compatibilidad de los datos en los ensayos y la estabilidad de los modelos basados en sólo unas pocas observaciones (el número de ensayos en el metaanálisis). Otra limitación, como en cualquier estudio de riesgo agregado, es la posibilidad de una «falacia ecológica», como se comentó en el capítulo 12.

Los diversos estudios en una revisión sistemática a veces abordan la efectividad de una serie de intervenciones relacionadas, no sólo una simple comparación entre la intervención y el grupo control. Por ejemplo, existen diferentes técnicas de cirugía bariátrica como la derivación yeyunoileal, la gastrectomía en manga, la banda gástrica ajustable, entre otras. Los ensayos aleatorizados han comparado varias combinaciones de estas entre sí y con la atención habitual, pero ningún estudio ha comparado cada técnica con todas las demás. El **metaanálisis en redes** es una forma de estimar la efectividad comparativa de las intervenciones que no se comparan directamente en los estudios reales, pero que pueden compararse de forma indirecta usando el modelo. Con un metaanálisis en redes, los investigadores fueron capaces de estimar los efectos respectivos de cada método de cirugía bariátrica en comparación con la atención habitual, demostrando que cada uno es efectivo e identificando la jerarquía de la efectividad (10).

METAANÁLISIS ACUMULATIVOS

Normalmente, los estudios de un diagrama de bosque se presentan por separado en orden alfabético por el primer autor o en orden cronológico. Otra forma de presentar la misma información es mediante un

booksmedicos.org

metaanálisis acumulativo. Los estudios que lo forman se ponen en orden cronológico, del más antiguo al más reciente, y se calcula un nuevo tamaño del efecto global y un IC cada vez que se cuente con los resultados de un nuevo estudio. De este modo, la figura representa un resumen del funcionamiento de todos los estudios hasta el momento de un nuevo ensayo. Este es un razonamiento bayesiano, como se describió en el capítulo 11, donde cada ensayo nuevo modifica la creencia anterior sobre la efectividad comparativa, establecido por los ensayos que se realizaron anteriormente.

El siguiente ejemplo ilustra el tipo de percepciones que un metaanálisis acumulativo puede proporcionar y también muestra cómo los metaanálisis en general y los análisis acumulativos en particular son útiles para establecer los efectos perjudiciales. Los ensayos individuales, que tienen el poder adecuado para detectar la efectividad, no tienen el poder suficiente para detectar los efectos perjudiciales porque la tasa de aparición es sustancialmente menor. La combinación de los datos puede acumular suficientes episodios para detectar los efectos perjudiciales.

Ejemplo

El rofecoxib es un fármaco antiinflamatorio no esteroideo (AINE) que tiene menos probabilidad de causar complicaciones gastrointestinales si se le compara con los AINE más antiguos. Los estudios de los pacientes que tomaban este fármaco informaron de un aumento en la tasa de episodios cardiovasculares, pero esto no se atribuyó al fármaco debido a que los intervalos de confianza eran amplios, no incluían el efecto y también porque el riesgo en el estudio más grande se explicó como un efecto protector del fármaco comparativo, el naproxeno. Los investigadores realizaron un metaanálisis de 16 ensayos aleatorizados de rofecoxib contra un control (placebo u otro AINE) (11). Todos los estudios eran pequeños a excepción de uno, e incluían 1-6 episodios cardiovasculares por ensayo, aunque hubo 24 episodios en el único ensayo grande. El riesgo relativo global de rofecoxib para el infarto de miocardio, incluyendo los datos del ensayo más grande, fue de 2,24 (IC del 95 % de 1,24-4,02). ¿Cuándo se volvió aparente el riesgo por primera vez? Un metaanálisis acumulativo (fig. 13-5) muestra que el efecto estadísticamente significativo se volvió aparente en el año 2000, sobre todo debido a la información aportada por el único estudio grande. Los estudios posteriores tendieron a consolidar este hallazgo y aumentar la precisión estadística. El rofecoxib fue retirado del mercado en 2004, 4 años después que un metaanálisis acumulativo hubiera mostrado que el riesgo cardiovascular estaba presente en los niveles convencionales de significancia estadística.

Los metaanálisis acumulativos se han utilizado para demostrar el

booksmedicos.org

momento en que la comunidad científica *podría haber sabido* sobre la efectividad o el daño de haber tenido un metaanálisis de la evidencia, aunque ahora es posible tener los resultados de un metaanálisis en tiempo real. En la Cochrane Collaboration, los metaanálisis se actualizan cada vez que se cuenta con un nuevo ensayo. Para cada metaanálisis nuevo, el tamaño del efecto global representa la acumulación de evidencia hasta el momento de la actualización, aunque sin demostrar expresamente los tamaños del efecto globales del pasado.

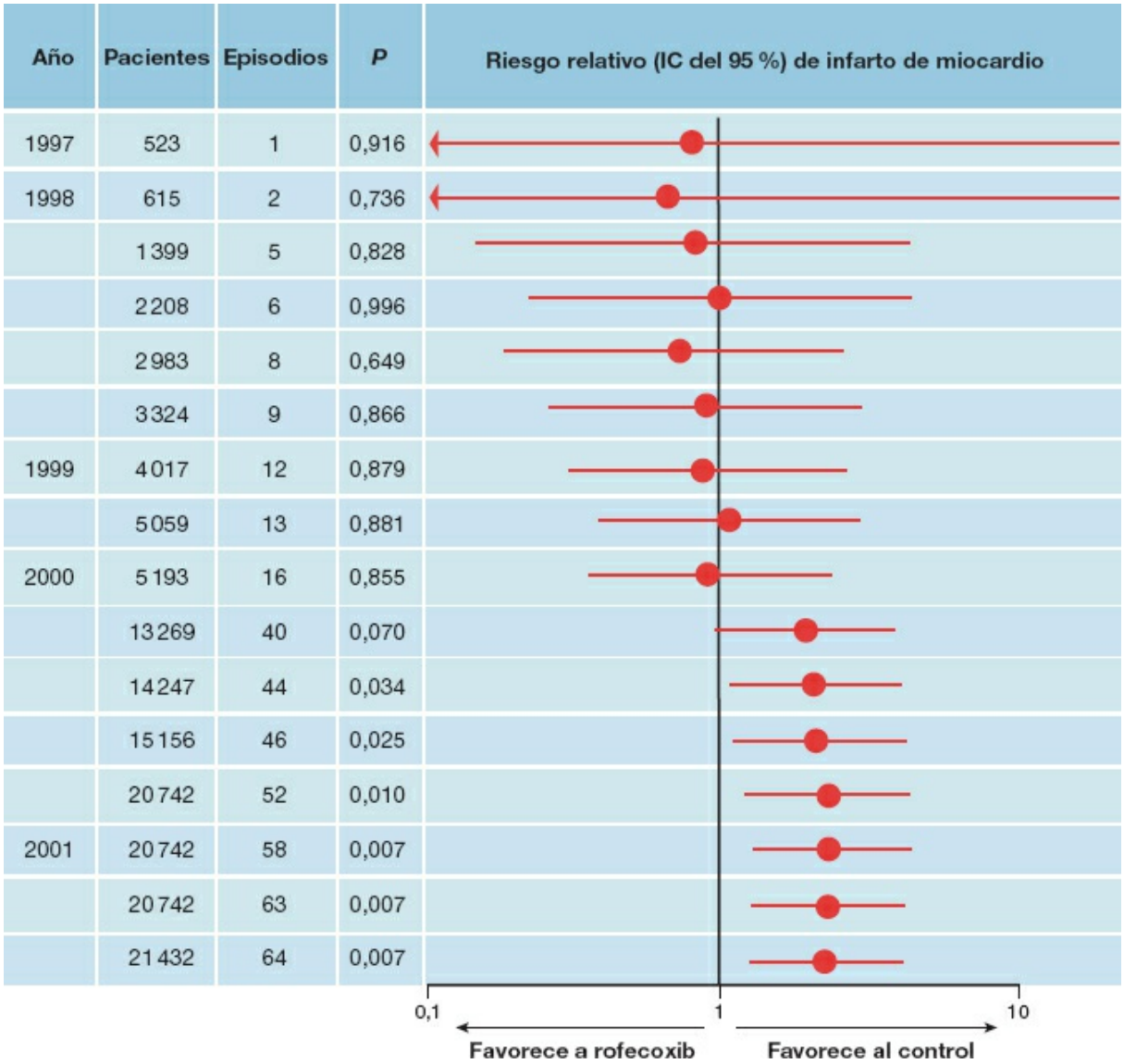


Figura 13-5 Un metaanálisis acumulativo sobre los estudios que comparan los efectos de rofecoxib con placebo sobre las tasas de infarto de miocardio. (Redibujado con autorización de Juni P, Nartey L, Reichenbach S, et al. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. Lancet 2004;364:2021-2029.)

booksmedicos.org

REVISIONES SISTEMÁTICAS DE ESTUDIOS OBSERVACIONALES Y PRUEBAS DIAGNÓSTICAS

Hemos comentado las revisiones sistemáticas y los metaanálisis utilizando los ensayos aleatorizados controlados como ejemplos. Sin embargo, las revisiones sistemáticas también son útiles para otro tipo de estudios, como se muestra en el siguiente resumen de estudios observacionales.

Ejemplo

Los pacientes con tromboembolismo venoso pueden presentar recurrencias después de descontinuar la anticoagulación. Algunos investigadores obtuvieron los datos individuales de 2 554 pacientes en siete estudios de cohorte que siguieron a los pacientes tras descontinuar la anticoagulación (12). Debido a que se contaba con la información de cada paciente, se estimó la incidencia con un análisis del tiempo transcurrido hasta el episodio. Los hombres tuvieron un mayor riesgo de tromboembolismo que las mujeres (HR = 2,2). Después de ajustar según el tromboembolismo inicial asociado a hormonas en las mujeres, se mantuvo el exceso de riesgo en los hombres. La tasa de recidiva no fue diferente entre los hombres y las mujeres cuyo primer episodio ocurrió después de la exposición a un factor de riesgo importante. Debido al mayor tamaño de la muestra obtenida de los resultados agrupados y por la disponibilidad de los datos de cada paciente, fue posible examinar la pregunta principal (si las tasas de recurrencia eran diferentes entre los hombres y las mujeres) y las preguntas secundarias (si los efectos eran diferentes en las mujeres menopáusicas que tomaron hormonas o si el episodio inicial ocurrió después de la exposición a un factor de riesgo importante).

El rendimiento de una prueba diagnóstica, al ser una pregunta relativamente específica, también se puede adaptar a las revisiones sistemáticas, como muestra el siguiente ejemplo.

Ejemplo

Los pacientes con dolor lumbar y en las piernas pueden tener un disco intervertebral herniado que comprime la raíz del nervio espinal. Los médicos de atención primaria dependen de la información obtenida en la historia clínica y la exploración física para seleccionar a los pacientes que podrían beneficiarse de los estudios de imagen y la cirugía. El dolor causado por la elevación de la pierna recta es una de las pruebas principales para el diagnóstico de un disco herniado. ¿Cuál es el desempeño de esta prueba? Los investigadores de la Cochrane Collaboration realizaron una revisión sistemática de los estudios que describen la sensibilidad y la especificidad de la elevación de la pierna recta (13). Encontraron 10 estudios que utilizaron los hallazgos quirúrgicos como patrón de referencia (fig. 13-6). La sensibilidad fue

booksmedicos.org

consistentemente alta, pero las especificidades correspondientes fueron bajas, calculadas de manera imprecisa y bastante variables. Debido a la heterogeneidad, sobre todo para la especificidad, los resultados no se agruparon. Los investigadores informaron que no hubo una única explicación para la heterogeneidad, pero sí una variación en la forma de interpretar la prueba y el estándar de referencia como anormal. Una curva ROC sugirió el equilibrio entre la sensibilidad y la especificidad en diferentes puntos de corte para una prueba anormal. La mayoría de los estudios fueron demasiado pequeños para apoyar una evaluación más completa sobre las causas de heterogeneidad.

FORTALEZAS Y DEBILIDADES DE LOS METAANÁLISIS

Los metaanálisis pueden hacer muchas contribuciones a las revisiones sistemáticas si están justificados por resultados relativamente homogéneos de los estudios que los forman. Pueden establecer que un efecto está presente o ausente con más autoridad que los ensayos individuales o las formas menos formales de recapitular los efectos. La combinación (agrupación) permite estimar los tamaños del efecto con mayor precisión para que los médicos puedan comprender mejor qué tan grande o pequeño podría ser el verdadero tamaño del efecto. Los metaanálisis pueden detectar las complicaciones del tratamiento o las diferencias de los efectos entre los subgrupos, problemas que no podrían ser abordados por ensayos individuales, que por lo general no cuentan con el poder estadístico. Los metaanálisis permiten reconocer el momento en el que se ha establecido la efectividad o el daño, mientras que esto es mucho más difícil con revisiones menos formales.

booksmedicos.org

Estudio	TP	FP	FN	TN	Sensibilidad	Especificidad
Albeck 1996	51	15	10	4		
Charnley 1951	63	8	11	6		
Demircan 2002	179	18	5	82		
Gurdjian 1961	929	12	22	13		
Hakelius 1972	1411	422	56	70		
Kerr 1988	98	20	2	16		
Knutsson 1961	155	18	7	2		
Kosteljanetz 1984	44	23	12	21		
Kosteljanetz 1988	40	6	5	1		
Spangfort 1972	2088	308	69	39		

Figura 13-6 Revisión sistemática de la realización de pruebas diagnósticas. La sensibilidad y la especificidad de la elevación de la pierna recta como prueba para detectar una hernia de disco lumbar en pacientes con dolor lumbar y dolor en la pierna. (Redibujado con autorización de van der Windt DA, Simons E, Riphagen II, et al. Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain. Cochrane Database Syst Rev 2010;[2]:CD007431.)

Las desventajas de los metaanálisis incluyen la tentación de agrupar estudios bastante diferentes, proporcionando una estimación engañosa sobre los efectos y apartando la atención del motivo por el que existen diferencias en los efectos. Los metaanálisis no incluyen información basada en la biología de la enfermedad, la experiencia clínica, y la aplicación práctica de la mejor evidencia en la atención del paciente, otras dimensiones de la atención que pueden tener una influencia tan grande en los resultados dirigidos al paciente como lo tiene la elección de los medicamentos o los procedimientos.

¿Qué se acerca más a la verdad, la mejor investigación individual o los metaanálisis? Es una elección falsa. Los metaanálisis no pueden ser mejores que la solidez científica de los estudios individuales que están resumiendo. Algunas veces la mayor parte del peso en un metaanálisis se deposita en un solo ensayo, por lo que hay relativamente poca diferencia entre la información del ensayo y en el resumen de todos los estudios sobre la misma pregunta. En su mayor parte, los dos –los estudios sólidos más grandes y los metaanálisis de todos los estudios relacionados– se

booksmedicos.org

complementan entre sí en lugar de competir. Cuando están en desacuerdo, el problema principal es el motivo por el que discrepan, que debe buscarse examinando los propios estudios, no los métodos en general.

Preguntas de revisión

Lea las siguientes preguntas y seleccione la mejor respuesta.

13-1. Una revisión sistemática de estudios observacionales de vitaminas antioxidantes para evitar las enfermedades cardiovasculares reúne los resultados de 12 estudios con el fin de obtener un resumen de la magnitud del efecto y un IC. ¿Cuál de las siguientes posibilidades sería la base lógica con más peso para combinar los resultados de los estudios?

A. Desea obtener una conclusión más generalizable.
B. Una prueba estadística revela que los estudios son heterogéneos.
C. La mayor parte de los estudios integrantes son estadísticamente significativos.
D. Los estudios integrantes tienen sesgos diferentes, y hasta cierto punto complementarios.
E. Usted puede obtener una estimación más precisa sobre el tamaño del efecto.

13-2. Debe realizar un juicio de una revisión de las publicaciones sobre si el alcohol supone un factor de riesgo para el cáncer de mama. El revisor ha buscado en MEDLINE y ha encontrado varios estudios observacionales sobre este tema, pero no ha buscado en otros sitios. Las siguientes posibilidades son limitaciones de esta estrategia de búsqueda *excepto:*

A. Los estudios con resultados negativos tienden a no publicarse.
B. Las búsquedas en MEDLINE normalmente omiten algunos artículos aunque estos estén incluidos en dicha base de datos.
C. MEDLINE no incluye todas las revistas del mundo.
D. Se pueden buscar en MEDLINE simultáneamente tanto el

booksmedicos.org

contenido como los métodos.

13-3. Una revisión sistemática de los antiagregantes plaquetarios para prevenir enfermedades cardiovasculares combina pacientes, no ensayos. ¿Cuál de las siguientes es una ventaja de este enfoque?

A. Es más eficaz para el investigador.
B. Es posible el análisis por subgrupos.
C. No es necesario seleccionar el modelo de efectos fijos o aleatorios cuando se combinan los datos.
D. Es menos probable que se produzcan sesgos de publicación.

13-4. ¿Cuál de los siguientes *no* se suele usar para definir una pregunta clínica específica estudiada mediante ensayos aleatorizados controlados?

A. Se toman en consideración las covariables.
B. Intervenciones (p. ej., exposición o tratamiento experimental).
C. Grupo comparativo (p. ej., los pacientes que toman placebo en un ensayo aleatorizado).
D. Resultados.
E. Pacientes en los ensayos.

13-5. ¿Cuál de los siguientes describe mejor los métodos para medir la calidad de un estudio?

A. La validez de las medidas de resumen sobre la calidad del estudio está bien establecida.
B. Una descripción de la calidad del estudio es una parte útil de las revisiones sistemáticas.
C. En una medida de resumen de la calidad, las fortalezas del estudio pueden compensar sus debilidades.
D. Las medidas individuales de calidad como la aleatorización y el enmascaramiento no se asocian con los resultados del estudio.

13-6. ¿Cuál de los siguientes tipos de estudio tienen la menor probabilidad de publicarse?

A. Estudios pequeños positivos.

booksmedicos.org

B. Estudios grandes negativos.
C. Estudios grandes positivos.
D. Estudios pequeños negativos.

13-7. ¿Cuál de las siguientes es una ventaja comparativa de las revisiones tradicionales («narrativas») sobre las revisiones sistemáticas?

A. Los lectores pueden confirmar que la evidencia citada ha sido seleccionada sin sesgos.
B. Las revisiones narrativas pueden revisar un gran número de temas relacionados con la atención de un trastorno.
C. Dependen de la experiencia y el juicio de un experto en la materia.
D. Proporcionan un resumen cuantitativo de los efectos.
E. La fortaleza científica de los estudios citados se evalúa explícitamente.

13-8. ¿Cuál de los siguientes factores no siempre forma parte de un diagrama de bosque típico?

A. El número de estudios que cumplió con estándares altos de calidad.
B. El tamaño del efecto global o combinado con el intervalo de confianza.
C. Las estimaciones puntuales del tamaño del efecto para cada estudio.
D. Los intervalos de confianza para cada estudio.
E. El tamaño o el peso aportado por cada estudio.

13-9. ¿Cuál de los siguientes no se puede utilizar para identificar la severidad de una enfermedad como una causa de heterogeneidad en una revisión sistemática con metaanálisis?

A. Controlar la gravedad de una enfermedad en un metaanálisis individual.
B. Un modelo matemático que relaciona la gravedad promedio de la enfermedad con el resultado de los estudios que lo forman.
C. Una comparación de los tamaños del efecto globales en los

booksmedicos.org

estudios, estratificados mediante la gravedad promedio de la enfermedad.

D. Un análisis por subgrupo de los datos individuales de los pacientes.

13-10. ¿Cuál de las siguientes es la mejor justificación para combinar los resultados de cinco estudios en un único efecto global?

A. Los pacientes, las intervenciones y los resultados son relativamente similares.

B. Todos los estudios son de alta calidad.

C. Una prueba estadística no detecta la heterogeneidad.

D. Se ha descartado el sesgo de publicación mediante un diagrama de embudo.

E. Usar el modelo de efectos aleatorios para calcular el tamaño del efecto global.

13-11. ¿Cuál de las siguientes es una ventaja del modelo de efectos aleatorios sobre el modelo de efecto fijo?

A. Se puede usar para estudios con análisis del tiempo transcurrido hasta el episodio.

B. Describe el tamaño del efecto global para una sola pregunta estrechamente definida.

C. Es más adecuado para los metaanálisis sobre el desempeño de una prueba diagnóstica.

D. Proporciona resultados más realistas cuando existe heterogeneidad entre los estudios.

E. Los intervalos de confianza tienden a ser más estrechos.

Encontrará las respuestas en el apéndice A.

BIBLIOGRAFÍA

1. Sackett DL, Richardson WS, Rosenberg W, et al. Evidence-based medicine. *How to practice and teach EBM*. New York; Churchill Livingstone; 1997.
2. Reichenbach S, Sterchi R, Scherer M, et al. Meta-analysis: chondroitin for osteoarthritis of the knee and hip. *Ann Intern Med* 2007;146:580–590.
3. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials* 1996;17:1–12.

booksmedicos.org

4. Balk EM, Bonis PAL, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;287:2973–2982.
5. Verdon F, Burnand B, Stubi CL, et al. Iron supplementation for unexplained fatigue in nonanemic women: double-blind, randomized placebo controlled trial. *BMJ* 2003;326:1124–1228.
6. El-Tawil S, Al Musa T, Valli H, et al. Quinine for muscle cramps. *Cochrane Database Syst Rev* 2010;(12):CD005044.
7. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ* 2001;322:1479–1480.
8. Bjelakovic G, Nikolova D, Simonetti RG, et al. Antioxidant supplements for prevention of gastrointestinal cancers: a systematic review and meta-analysis. *Lancet* 2004;364:1219–1228.
9. Askie LM, Ballard RA, Cutter GR, et al. Inhaled nitric oxide in preterm infants: an individual-level data meta-analysis of randomized trials. *Pediatrics* 2011;128:729–739.
10. Padwal R, Klarenbach S, Wiebe N, et al. Bariatric surgery: a systematic review and network meta-analysis of randomized trials. Obes Rev 2011;12:602–621.
11. Juni P, Nartey L, Reichenbach S, et al. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *Lancet* 2004;364:2021–2029.
12. Douketis J, Tosetto A, Marcucci M, et al. Risk of recurrence after venous thromboembolism in men and women: patient level meta-analysis. *BMJ* 2011;342:d813.
13. van der Windt DA, Simons E, Riphagen II, et al. Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain. *Cochrane Database Syst Rev* 2010;(2):CD007431.

booksmedicos.org

Capítulo 14

Administración del conocimiento

¿Dónde está la sabiduría que hemos perdido en el conocimiento? ¿Dónde está el conocimiento que hemos perdido en la información?

—T.S. Eliot
1934

PALABRAS CLAVE

Gestión o administración del conocimiento
Conflicto de intereses
Fraude científico
Lugar de atención sanitaria
Aprendizaje oportuno
Guías de práctica clínica
MEDLINE
PubMed
EMBASE
Revisión crítica
Resúmenes estructurados

Buscar la respuesta más apropiada para una cuestión clínica específica es como buscar una aguja en un pajar. La información básica está mezclada con una gran cantidad de «supuestos» y opiniones menos verosímiles, y separar el trigo de la paja no es tarea fácil. Sin embargo, es lo que los médicos han de realizar. La lectura crítica tiene sólo la misma utilidad que la información que se encuentra.

booksmedicos.org

La **gestión o administración del conocimiento** consiste en organizarlo de forma efectiva y eficaz. Esto era un trabajo complicado en la época en que sólo había prensa escrita. Por fortuna, la administración del conocimiento se ha vuelto mucho más sencilla en la era de la información electrónica. Hay más estudios y mejores sobre un campo muy abundante de temas clínicos, amplio acceso a los resultados de la investigación y métodos eficientes para clasificar rápidamente los artículos por tema y validez científica. Estas oportunidades son consecuencia de la generalización del uso de las computadoras, Internet y las bases de datos electrónicas con fines clínicos.

Buscar información puede parecer que no sea una de las principales prioridades de los médicos que todavía se están formando. Están rodeados de información, mucha más de la que pueden manejar con comodidad, y cuentan con innumerables especialistas que les ayudan a decidir lo que deben tomar en serio y lo que deben ignorar. Sin embargo, desarrollar el propio proyecto para administrar el conocimiento después es fundamental, ya sea en la práctica o en la teoría. Incluso con las últimas novedades, la administración efectiva y eficaz de los conocimientos es una tarea que exige mucho esfuerzo. En este capítulo analizamos los enfoques modernos para la gestión de conocimiento. Discutiremos las cuatro tareas básicas: buscar la información, mantenerse al día con los nuevos avances en su área, mantenerse conectado con la medicina como una profesión y ayudar a los pacientes a encontrar buena información sobre la salud por sí mismos.

PRINCIPIOS BÁSICOS

Hay varios aspectos de la administración del conocimiento que abarcan todas las actividades.

¿Lo hace uno mismo o delega el trabajo?

En primer lugar, los médicos deben preguntarse: «¿Voy a buscar y a evaluar los resultados de la investigación clínica yo mismo o delego este trabajo a otra persona?» La respuesta es ambas. Los médicos deben ser capaces de encontrar y criticar la información por su cuenta; es una habilidad básica en la medicina clínica. No obstante, desde el punto de vista práctico, no es posible hacer la tarea en solitario, teniendo en cuenta toda la información que se necesita. Hay demasiadas cuestiones en un día y muy poco tiempo para responderlas por su cuenta. Por lo tanto, deben encontrar personal de

booksmedicos.org

confianza que les ayude a administrar el conocimiento.

¿Qué tipo de recursos?

Se puede obtener información a través de una gran variedad de recursos, que van desde los libros impresos y las revistas hasta la información digital en internet, a la que se puede acceder mediante plataformas fijas y portátiles. Existen cintas de audio, cintas de vídeo y más. La información no es ni más ni menos sólida por la manera en que llega hasta usted. La validez depende de los autores, los revisores y los editores, no del medio. Con todo, la disposición de varios soportes, con ventajas e inconvenientes que se complementan, facilita la búsqueda según las preferencias personales.

Un plan moderno para administrar el conocimiento debe basarse en la información electrónica de internet. La información fundamental para la medicina clínica está cambiando demasiado rápido como para que sólo los medios escritos sean suficientes. Por ejemplo, los descubrimientos clínicamente importantes sobre la terapia antiviral para el VIH, las tecnologías innovadoras de escaneo y la quimioterapia contra el cáncer basada en los avances científicos emergen año a año, incluso mes a mes. Internet puede seguir el ritmo con este cambio tan rápido y también complementar, pero no reemplazar, a las fuentes tradicionales.

Clasificación de la información

La clasificación permite que los médicos comprendan el valor básico de la información en cuestión de segundos. En general, la calidad de la evidencia (confianza en la estimación de los efectos) y la fuerza de las recomendaciones se clasifican por separado. La tabla 14-1 muestra un ejemplo de un esquema de clasificación ampliamente utilizado, denominado GRADE, con un principio similar a otros métodos de uso general. Esta clasificación es para las intervenciones; la clasificación de otro tipo de información está menos desarrollada. Considere que las recomendaciones se basan en la fortaleza de la evidencia obtenida en la investigación, dependen del equilibrio entre los beneficios y daños, y varían en qué tan enérgica y ampliamente se debe ofrecer la intervención a los pacientes. Aunque los criterios de clasificación son explícitos, la asignación de una clase depende en parte del juicio.

Informes engañosos sobre los hallazgos de investigación

booksmedicos.org

Hasta ahora se ha actuado como si las únicas amenazas a la validez de la investigación clínica publicada se originaran de las dificultades para aplicar buenos principios científicos al estudio de la enfermedad humana. Es decir, la validez depende del manejo de los sesgos y el azar. Por desgracia, hay otras amenazas a la validez de los resultados de investigación relacionadas con los propios investigadores y el ambiente social, político y económico en el que trabajan. Nos referimos a la tendencia muy humana de informar de los resultados de investigación de acuerdo con el interés personal en los resultados.

El **conflicto de intereses** existe cuando los intereses privados de los investigadores compiten con su responsabilidad de ser investigadores sin sesgos. Hay muchos conflictos de intereses:

- Conflicto de intereses financieros: cuando el ingreso personal o familiar está relacionado con los resultados de la investigación (suele considerarse que este conflicto es el más poderoso y el más difícil de detectar).
- Relaciones personales: apoyar a los amigos y abatir a los rivales.
- Pasión intelectual: estar a favor de las ideas propias y en contra de las contrarias.
- Lealtad institucional: poner los intereses de la propia escuela, compañía u organización por encima de los demás.
- Promoción profesional: los investigadores obtienen más crédito académico por publicar resultados interesantes en revistas de élite.

El conflicto de intereses existe en relación con un tema específico, no en general, sin importar si realmente ha modificado el comportamiento.

¿Cómo se expresa el conflicto de intereses? El **fraude científico,** estafa, fabricación y plagio, es un ejemplo extremo. El informe selectivo de los resultados es menos extremo, ya sea por no informar de los resultados desagradables (sesgo de publicación) o informar de los resultados en función de si parecen ser los «correctos». En ocasiones, la industria que patrocina la investigación puede bloquear la publicación o alterar la forma en que se informa de los resultados. Para crear un registro público con el fin de determinar si esto ha ocurrido, actualmente se registran los ensayos aleatorizados controlados en sitios de internet a disposición del público antes de la recolección de datos, lo que permite seguir de cerca si se han publicado los resultados cuando se esperaba y si los criterios de valoración fueron los

booksmedicos.org

mismos que cuando comenzó el ensayo (1).

Los esfuerzos para «girar» los resultados por la forma en que se describen son más sutiles y más difíciles de detectar, por ejemplo, dando a entender que un valor muy bajo de *P* significa que los resultados son clínicamente importantes o describiendo los resultados como «grandes» cuando la mayoría de nosotros pensaría que no lo son (2). Todos dependemos de revisores y editores para restringir lo peor de este tipo de desviación editorial en los artículos científicos.

Mencionamos estas influencias un tanto sórdidas en la información en la que dependen los médicos (y sus pacientes) porque son, en algunas situaciones, tan reales e importantes como la aplicación bien informada de los intervalos de confianza y el control de la confusión, el dominio habitual de la epidemiología clínica. La investigación y su interpretación son actividades humanas y, por lo tanto, siempre estarán matizadas en algún grado, con resultados egoístas. Se están realizando esfuerzos para limitar los sesgos relacionados con los conflictos de intereses, sobre todo al insistir en la revelación completa, pero también al excluir a personas con evidente conflicto de intereses debido a la revisión por colegas de sus manuscritos y subvenciones, la autoría de artículos de revisión y editoriales, y pertenecer a un grupo que elabora guías.

booksmedicos.org

Tabla 14-1

Clasificación de las recomendaciones para el tratamiento de acuerdo con la calidad de la evidencia (confianza en la estimación del efecto, A-C) y fuerza de la recomendación (1-2) con sus implicaciones. Basado en las guías GRADE

Grado de recomendación	Claridad del riesgo/ beneficio	Calidad de la evidencia que lo apoya	Implicaciones
1A. Recomendación fuerte, evidencia de alta calidad	Los beneficios claramente superan a los riesgos y las cargas, o viceversa	Evidencia consistente de ensayos aleatorizados controlados bien ejecutados, o evidencia abrumadora de alguna otra forma. Es improbable que la investigación adicional cambie la confianza en las estimaciones sobre los beneficios y los riesgos	Recomendaciones sólidas aplicables para la mayoría de los pacientes en la mayoría de las circunstancias sin reservas. Los médicos deben seguir una recomendación fuerte a menos que exista una justificación clara y contundente para un enfoque alternativo
1B. Recomendación fuerte, evidencia de moderada calidad	Los beneficios claramente superan a los riesgos y las cargas, o viceversa	Evidencia de ensayos aleatorizados controlados con limitaciones importantes (resultados inconsistentes, fallos metodológicos o imprecisión), o evidencia muy sólida de alguna otra forma. Es probable que la investigación adicional (si se lleva a cabo) cambie nuestra confianza en las estimaciones sobre los beneficios y los riesgos	Recomendaciones fuertes aplicables para la mayoría de los pacientes. Los médicos deben seguir una recomendación fuerte a menos que exista una justificación clara y contundente para un enfoque alternativo
1C. Recomendación fuerte, evidencia de baja calidad	Los beneficios parecen superar a los riesgos y las cargas, o viceversa	Evidencia de estudios observacionales, experiencia clínica no sistemática, o ensayos aleatorizados controlados con fallos graves. Cualquier estimación sobre el efecto es incierta	Recomendación fuerte aplicable para la mayoría de los pacientes. Una parte de la evidencia que apoya la recomendación es de baja calidad
2A. Recomendación débil, evidencia de alta calidad	Los beneficios casi están equilibrados con los riesgos y las cargas	Evidencia consistente de ensayos aleatorizados controlados bien ejecutados, o evidencia abrumadora de alguna otra forma. Es improbable que la investigación adicional cambie la confianza en las estimaciones de los beneficios y los riesgos	Recomendación débil. La mejor acción puede diferir dependiendo de las circunstancias o los valores del paciente o la sociedad
2B. Recomendación débil, evidencia de moderada calidad	Los beneficios casi están equilibrados con los riesgos y las cargas, con cierto grado de incertidumbre en las estimaciones de los beneficios, riesgos y cargas	Evidencia de ensayos aleatorizados controlados con limitaciones importantes (resultados inconsistentes, fallos metodológicos o imprecisión), o evidencia muy sólida de alguna otra forma. Es probable que la investigación adicional (si se lleva a cabo) cambie nuestra confianza en las estimaciones sobre los beneficios y los riesgos	Recomendación débil. Es probable que los enfoques alternativos sean mejores para algunos pacientes bajo algunas circunstancias
2C. Recomendación débil, evidencia de baja calidad	Incertidumbre en las estimaciones de los beneficios, riesgos y cargas; los beneficios casi están equilibrados con los riesgos y las cargas	Evidencia de estudios observacionales, experiencia clínica no sistemática o ensayos aleatorizados controlados con fallos graves. Cualquier estimación sobre el efecto es incierta	Recomendación muy débil. Otras alternativas pueden ser igualmente razonables

Adaptado de Guyatt GH, Oxman AD, Vist GE, et al. for the GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924-926.

BÚSQUEDA DE RESPUESTAS A CUESTIONES CLÍNICAS

booksmedicos.org

Los médicos deben ser capaces de buscar las respuestas a las preguntas que surgen durante la atención de sus pacientes. Esto es necesario para las cosas que desconocen, pero también para revisar los hechos que creen que conocen, pero que puede que no sea así, porque el fundamento de la información asistencial del paciente cambia constantemente.

Lo mejor es obtener las respuestas a las preguntas justo en el momento y en el lugar en que se presenten durante el cuidado de los pacientes. A esto se le ha llamado el **lugar de atención médica** y al aprendizaje asociado se denomina **aprendizaje oportuno** (justo a tiempo). Las respuestas pueden ser usadas para dirigir la toma de decisiones clínicas para el paciente en cuestión. Además, es más probable recordar lo que se aprende que la información encontrada fuera de contexto en un aula, sala de conferencias, libro o revista, apartada de la necesidad de conocer un paciente específico. En cualquier caso, posponer las respuestas a las preguntas para un momento posterior a menudo significa que no serán respondidas en absoluto.

Para que se lleve a cabo el aprendizaje oportuno se debe cumplir con varias condiciones (tabla 14-2). La mayoría de los sitios de atención están sometidos a la presión del tiempo, así que se debe obtener rápidamente la respuesta. Como un pediatra dedicado a la consulta señaló: «¡Si añadiera sólo 1-2 min a cada consulta de un paciente, llegaría a casa 1 h más tarde al final del día!» Lo que necesitan los médicos no es *una* respuesta sino la mejor respuesta disponible de acuerdo con el conocimiento en ese momento. Necesitan conseguir información que se corresponda lo máximo posible con la situación clínica concreta que están tratando. Si un paciente es de avanzada edad y tiene varias enfermedades, la información sobre la investigación debe ser de pacientes ancianos con comorbilidades. Los médicos necesitan fuentes de información que puedan llevar con ellos mientras se desplazan por su trabajo de la consulta a casa (donde se les puede contactar por la noche y el fin de semana) y a hospitales y residencias de ancianos.

Cuando todo esto ocurre, y sin duda puede ocurrir, los resultados son extraordinariamente poderosos.

Tabla 14-2

Condiciones bajo las cuales está disponible la información en el lugar de atención médica

Condición	Fundamento

booksmedicos.org

Acceso rápido	La información debe estar disponible en cuestión de minutos para que se adapte en el bullicioso flujo de trabajo de la mayoría de los sitios para la atención de los pacientes
Actual	Debido a que la mejor información en la que se basan las decisiones clínicas está cambiando continuamente, suele ser necesario que la información sea electrónica (como una cuestión práctica, de internet)
Adaptada al tema específico	Los médicos necesitan información que coincida lo más posible con la situación real de su propio paciente
Clasificada por su validez científica	Hay una gran cantidad de información para casi cualquier pregunta clínica, pero sólo una pequeña proporción de la misma es científicamente sólida y clínicamente relevante
Disponible en situaciones clínicas	Los médicos no pueden abandonar su lugar de trabajo para buscar las respuestas; deben encontrarlas en el sitio donde trabajan

Ejemplo

Un paciente acude a su consulta porque va a viajar a Ghana y quiere que le informe sobre la prevención del paludismo. Usted es consciente de que la susceptibilidad del parásito del paludismo a los fármacos antipalúdicos varía en todo el mundo y cambia constantemente. Los Centers for Disease Control and Prevention tienen una página web (http://www.cdc.gov) con información actualizada para las personas que han de viajar a cualquier parte del mundo. Usando la computadora en su clínica, averigua rápidamente qué fármaco profiláctico debe tomar este paciente y durante cuánto tiempo, antes, durante y después del viaje. También se informa de que la persona se debe vacunar contra la hepatitis A y B, la tifoidea y la fiebre amarilla. El sitio enumera las clínicas que tienen estas vacunas y también muestra que el norte de Ghana, que visitará su paciente, está en la «franja de la meningitis», por lo que también debe vacunarse contra el meningococo. Esta información es una síntesis actualizada sobre los mejores consejos del mundo y está a su disposición en unos segundos.

Soluciones

Colegas clínicos

Una red de colegas con experiencia variada y complementaria es una manera tradicional de obtener información en el lugar de atención médica. Muchos médicos han identificado a los líderes de opinión locales para este propósito. Por supuesto, estos líderes de opinión deben tener sus propias fuentes de información, presumiblemente mejores que las de otros colegas.

Libros de texto electrónicos

Los libros de texto, incluso bibliotecas completas, se encuentran en internet, y las escuelas de medicina, los sistemas sanitarios y las sociedades profesionales las han puesto a disposición de los médicos. Por ejemplo, UpToDate (http://www.uptodate.com), una fuente de información electrónica dirigida a los médicos, es el producto de miles de médicos autores y editores que revisan 9 000 temas equivalentes a 90 000 páginas impresas (si es que alguna vez se imprimen)[†]. La información se actualiza continuamente, es revisada por colegas, se puede buscar y está vinculada a resúmenes de la investigación original, y sus recomendaciones están clasificadas. UpToDate está disponible en el lugar de la atención médica en todo el mundo, siempre que se pueda acceder a internet mediante computadoras o plataformas móviles.

Ejemplo

Un autor estaba tratando a pacientes en Boston durante la amenaza del carbunco en 2001. En esos tiempos, los terroristas biológicos diseminaron las esporas de carbunco a través del sistema postal de EE.UU., provocando docenas de casos y cinco muertes. Una mujer joven acudió a urgencias porque estaba preocupada por si una lesión que se hizo recientemente en la piel estuviera causada por carbunco. Cuando se le dijo que no era así, ella respondió (con cierta descortesía) «¿Cómo lo sabe si nunca lo ha visto?». A lo que el doctor respondió, «Por supuesto que no, ninguno de nosotros lo ha hecho, pero sabemos cómo se ve. Venga conmigo y le mostraré». Utilizando UpToDate en una computadora de la consulta fue capaz de mostrarle varias fotografías de las lesiones dermatológicas debidas a carbunco, que eran bastante diferentes de las suyas, así que ella confió en que no era en realidad la próxima víctima del carbunco.

Otros libros de texto, como *ACP Medicine, Harrison* en línea, y muchos otros de subespecialidades también están disponibles en formato electrónico.

Guías de práctica clínica

booksmedicos.org

Las **guías de práctica clínica** son consejos para los médicos sobre la asistencia de los pacientes con determinadas afecciones. Además de dar recomendaciones, las buenas guías explicitan la base de la evidencia y los fundamentos de estas recomendaciones.

Al igual que la medicina basada en la evidencia, las guías pretenden ser un punto de inicio para que las decisiones sobre pacientes concretos se puedan modificar con criterios clínicos; es decir, son guías, no reglas. Las guías de alta calidad representan la aplicación acertada de las pruebas de la investigación a las realidades de la asistencia clínica. No obstante, la calidad de estas instrucciones varía. En la tabla 14-3 hay un resumen de algunos criterios para guías verosímiles creadas por el U.S. Institute of Medicine. En la *National Guideline Clearinghouse* en línea, http://www.guidelines.gov, se puede encontrar una lista relativamente exhaustiva de guías.

Tabla 14-3

Normas para unas guías de práctica clínica fiables

Estándar	Explicación
Transparencia	Se ha expresado cómo se desarrolló y se financió la guía y está accesible al público
Conflicto de intereses	Se han dado a conocer los conflictos de intereses de los miembros del grupo relacionados con actividades financieras, intelectuales, institucionales y de los pacientes/público que influyen en la guía
Composición del grupo	La afiliación al grupo fue multidisciplinaria y equilibrada, incluyendo a una variedad de expertos en metodología y médicos, y a las poblaciones que se espera sean afectadas por la guía
Revisión sistemática	Las recomendaciones se basan en revisiones sistemáticas que cumplen elevados estándares de calidad
Evidencia y fuerza de la recomendación	Cada recomendación se acompaña de una explicación de su razonamiento subyacente, el nivel de confianza en la evidencia y la fuerza de la recomendación
Descripción de las	La guía establece precisamente la acción

booksmedicos.org

recomendaciones	recomendada y bajo qué circunstancias se debe realizar
Revisión externa	La guía se ha revisado por un espectro completo de las partes interesadas involucradas (p. ej., expertos científicos y clínicos, organizaciones y pacientes)
Actualización	La guía indica la fecha de la publicación y la revisión de la evidencia, además de los planes de actualización cuando exista evidencia nueva que podría cambiar sustancialmente la guía

Modificado de Institute of Medicine. Clinical Practice Guidelines We Can Trust. Washington, DC: National Academies Press; 2011. The standards were for developing guidelines and have been modified to guide users in recognize guidelines they can trust.

[†]Robert y Suzanne Fletcher son algunos de los cientos de editores de UpToDate.

La biblioteca Cochrane

Científicos clínicos de todo el mundo se han ofrecido como voluntarios para revisar las publicaciones internacionales sobre cuestiones clínicas específicas, sintetizarlas y almacenarlas en el sitio central, y mantenerlas actualizadas. La recopilación de las revisiones está disponible en http://www.cochrane.org. La base de datos Cochrane está incompleta, dado el gran número de cuestiones que se pueden tratar, aunque es una fuente excelente de información rigurosamente seleccionada de revisiones sistemáticas, con metaanálisis cuando están justificados, efectos de las intervenciones y, más recientemente, el cumplimiento de una prueba diagnóstica.

Bases de datos (PubMed y otras)

MEDLINE es una base de datos bibliográfica recopilada por la *U.S. National Library of Medicine*, que abarca unos 5 000 artículos de biomedicina y salud, la mayoría de ellos publicados en inglés. Está disponible de forma gratuita utilizando un motor de búsqueda, generalmente **PubMed** (http://www.ncbi.nlm.nih.gov/pubmed). MEDLINE se puede buscar en PubMed de acuerdo con el tema, revista, autor, año y diseño del estudio. Además de las citas, están disponibles algunos resúmenes. También se utiliza **EMBASE** (http://embase.com), que complementa lo que se encuentra en MEDLINE; además de estas dos hay muchas otras bases de datos bibliográficas con fines más especializados.

booksmedicos.org

Las búsquedas en PubMed están limitadas por dos tipos de errores de clasificación. En primer lugar, producen resultados falsos negativos; es decir, pasan por alto los artículos que realmente se desean. En segundo lugar, las búsquedas producen muchos resultados falsos positivos, es decir, encuentran más citas que las que realmente se desean de acuerdo con su validez científica y relevancia clínica. Por ejemplo, cuando se pidió a nefrólogos canadienses que utilizaran PubMed para responder a preguntas clínicas exclusivas de su área, fueron capaces de recuperar el 46% de los artículos relevantes y el cociente de artículos relevantes y no relevantes fue de 1/16 (3). Ambos problemas pueden reducirse pero no superarse por completo mediante mejores técnicas de búsqueda.

Las búsquedas en PubMed son fundamentales para los investigadores y los educadores que disponen de tiempo para llevar a cabo búsquedas minuciosas y clasificar los artículos que se obtienen, pero estas búsquedas son demasiado ineficaces y no son muy prácticas para ayudar a los médicos, sobre todo cuando necesitan responder rápido. Sin embargo, como una fuente más de información médica, PubMed es particularmente útil, sobre todo para saber si hay constancia de episodios poco frecuentes.

Ejemplo

Al visitar a un paciente que presenta dolor abdominal, usted cree que tiene linforreticulosis benigna (enfermedad por rasguño de gato). Tras descartar otras causas, se pregunta si el dolor abdominal puede deberse a linfadenopatía abdominal. Para saber si esto se ha comunicado alguna vez, realiza una búsqueda en PubMed y encuentra un informe de un caso sobre la fiebre por rasguño de gato y linfadenopatía abdominal (4). Con esta información, aunque sólo haya el informe de un caso, se siente algo más seguro con su diagnóstico y tratamiento.

Otros recursos en internet

Una gran cantidad de información médica se publica en internet, alguna es muy útil para los profesionales de la salud. Se puede encontrar a través de un motor de búsqueda como Google o Google Escolar y en sitios subencionados por el gobierno de EE.UU. como MedlinePlus (http://www.nlm.nih.gov/medlineplus) y HealthFinder (http://healthfinder.gov) para la información médica y Health Hotlines (http://healthhotlines.nlm.nih.gov) para obtener la información de contacto de

booksmedicos.org

las organizaciones de salud relacionadas. Otros países tienen sus propios recursos en internet.

OBSERVACIÓN DE PROGRESOS

Mantenerse al día de los progresos en una especialidad de la medicina no es una tarea fácil. No es que el ritmo de los descubrimientos que cambian la práctica sea incontrolable, sino que la información relevante está muy dispersa en muchas revistas y se mezcla con un gran número de artículos menos importantes.

Ejemplo

¿Están muy dispersos en las revistas los mejores artículos de una especialidad? Los editores de *ACP Journal Club* revisan con regularidad más de 100 revistas y seleccionan artículos de medicina interna relevantes desde el punto de vista científico y clínico para su publicación cada mes. Los resultados se publican en una revista bimensual. Este proceso proporciona una oportunidad para calcular, por lo menos en el caso de la medicina interna, el grado en que los artículos clave están dispersos en las revistas. En la figura 14-4 se muestra la proporción de artículos clave con los que un lector se encontrará según el número de revistas que lea, empezando por la revista más importante y continuando con otras de menor relevancia. Para encontrar la mitad de estos artículos, se tienen que leer con frecuencia 4 revistas, 8 para acceder al 75% de ellos y 20 artículos para encontrar los artículos clave en medicina interna.

booksmedicos.org

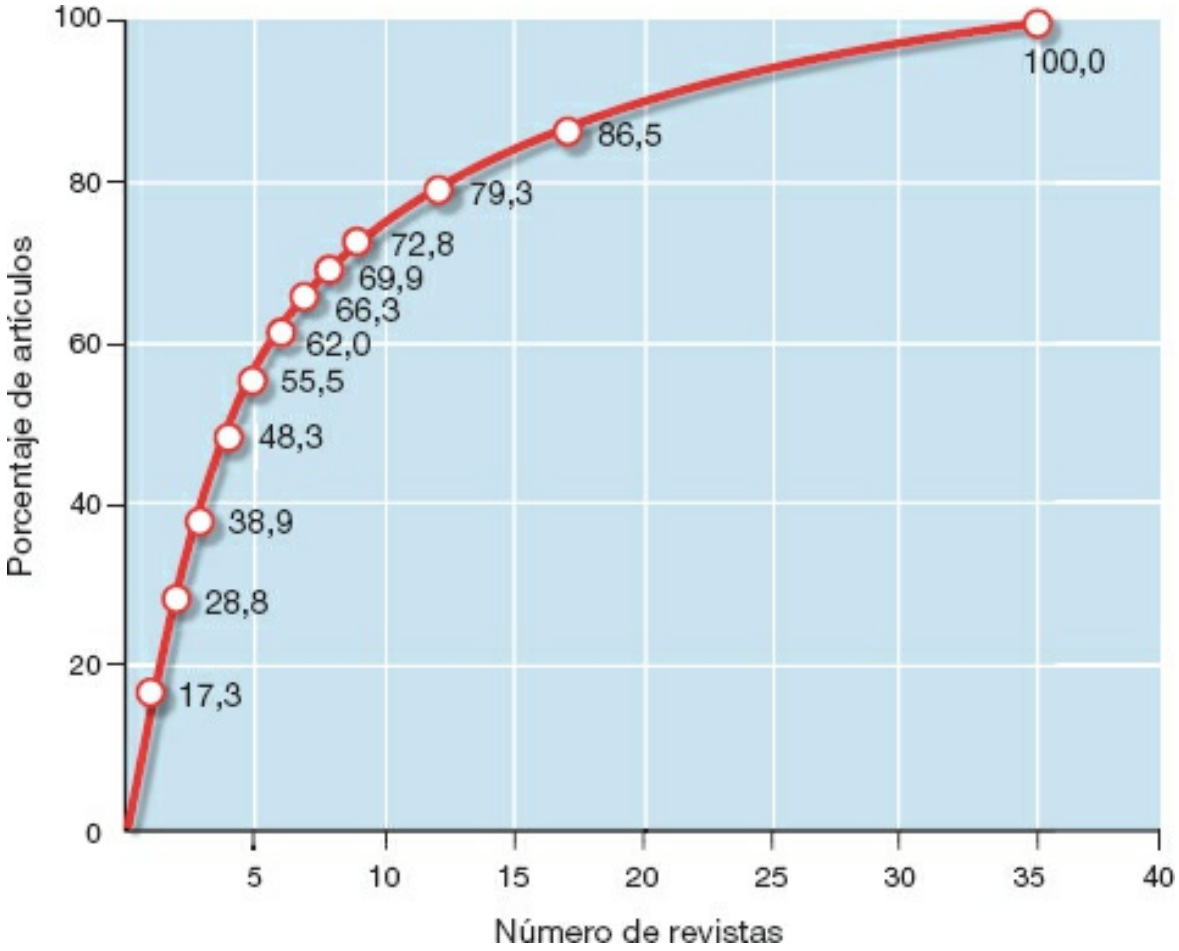


Figura 14-1 ¿Cuántas revistas tendrá que leer para estar al día con las publicaciones de su especialidad? Proporción de artículos científicamente relevantes de medicina interna según el número de revistas clasificadas de mayor a menor importancia. (De ACP Journal Club, 2011.)

Por lo tanto, no es posible que los lectores particulares, incluso haciendo un gran esfuerzo, encuentren todos los artículos esenciales de una especialidad. Deben delegar el trabajo a un intermediario de confianza, alguien que revise muchas revistas y seleccione los artículos en función de los criterios que se acuerden.

Afortunadamente, se puede disponer de ayuda. La mayoría de las especialidades clínicas patrocinan publicaciones que resumen artículos relevantes de su especialidad. Estas publicaciones varían en función de lo explícito y riguroso que sea el proceso de selección. Por un lado, el *ACP Journal Club* publica sus criterios para cada tipo de artículo (estudios de prevención, tratamiento, diagnóstico, pronóstico, entre otros) y proporciona una crítica de cada artículo que selecciona. Por otro lado, muchos boletines informativos incluyen resúmenes de artículos sin explicitar cómo se

booksmedicos.org

seleccionaron, ni cuáles son sus puntos fuertes y limitaciones.

En la actualidad hay varias maneras de obtener información nueva, los artículos de investigación publicados, las guías de práctica clínica, los documentos técnicos y los artículos de noticias, en el área específica de interés para que le sea enviada. Una forma es identificar temas específicos y hacer que la información nueva sobre estos le sea enviada en el momento en que surge, a través de una fuente RSS y de otros servicios. Otra es participar en el número cada vez mayor de redes sociales, como Face-book y blogs, en las que otras personas descubren y seleccionan información sobre la que usted podría querer saber y se la envían, al igual que usted les envía información nueva. Los ejemplos menos estructurados son los equipos de investigación que comparten artículos e historias novedosos relacionadas con su trabajo y los equipos de trabajo en los hospitales de enseñanza donde residentes, estudiantes y médicos encargados comparten artículos sobre los problemas médicos de sus pacientes. Las redes sociales pueden ser efectivas y eficientes si elige a los colegas adecuados para participar con ellos.

REVISTAS

Las revistas desempeñan un papel muy especial en las profesiones médicas. Todo lo que hemos dicho sobre la epidemiología clínica y la administración del conocimiento está sesgado en función de la investigación original publicada en revistas revisadas por colegas.

Los informes de las investigaciones se seleccionan y mejoran antes de su publicación mediante un proceso riguroso que integra la **revisión crítica,** los comentarios de los expertos en el tema que trata el artículo y los métodos que proporcionan asesoramiento sobre la conveniencia de publicar y cómo mejorar el manuscrito (el plazo del artículo antes de ser publicado). Los revisores son asesores del editor (o el equipo editorial), no quienes directamente deciden el destino del manuscrito. La revisión por colegas y la edición, además de la base de la evidencia y su justificación, están resumidos en el sitio oficial de la World Association of Medical Editors (http://www.wame.org) y el International Committee of Medical Journal Editors (http://www.icmje.org). Se ha demostrado que estas prácticas mejoran los manuscritos, pero los artículos publicados están lejos de ser perfectos (5); por lo tanto, los lectores deben estar agradecidos por los esfuerzos que hacen las revistas para mejorar los artículos, pero también

booksmedicos.org

deben mantener un cierto escepticismo en cuanto a la calidad del resultado final.

Los grupos de trabajo han definido la información que debe estar en un artículo de investigación completo de acuerdo con el tipo de estudio (ensayo aleatorizado controlado, evaluación de una prueba diagnóstica, revisión sistemática, etc.) (tabla 14-4). Los lectores pueden utilizar estas listas de cotejo para determinar si toda la información necesaria está incluida en un artículo, al igual que los investigadores los usan para asegurarse que sus artículos están completos.

Las revistas, por sí mismas, no son particularmente de ayuda de algunos elementos de gestión del conocimiento. Su lectura individual no es una forma de mantenerse al día con los nuevos avances científicos del sector o para buscar las respuestas a las preguntas clínicas. Son una vía para mantenerse al día en las grandes dimensiones del entorno clínico. Las opiniones, las historias, las hipótesis sin comprobar, los comentarios sobre artículos publicados, las declaraciones de valores profesionales, así como las descripciones del contexto histórico, social y político de la medicina actual, etc., reflejan la naturaleza compleja de la profesión (tabla 14-5). La riqueza de esta información completa la imagen clínica de muchos lectores. Por ejemplo, cuando *Annals of Internal Medicine* empezó con la publicación de historias sobre la profesión de médico (6), muchos lectores comentaron que aunque los informes de las investigaciones y revisiones eran fundamentales, las experiencias de los médicos era lo que más les interesaba.

Tabla 14-4

Guías para informar de los estudios de investigación

Tipo de estudio	Nombre de la declaración	Cita
Ensayos aleatorizados controlados	*Consolidated Standards of Reporting Trials*(CONSORT)	http://www.consort-statement.org
Pruebas diagnósticas	*Standards for Reporting of Diagnostic Test Accuracy* (STARD)	http://www.stard.org

booksmedicos.org

Estudios observacionales	*Strengthening the Reporting of Observational Studies in Epidemiology* (STROBE)	http://www.strobe-statement.org
Estudios no aleatorizados sobre intervenciones educativas, conductuales y de salud pública	*Transparent Reporting of Evaluations with Nonrandomized Design* (TREND)	http://www.cdc.gov/trendstatemen
Metaanálisis de ensayos aleatorizados controlados	*Quality of Reporting of Meta-analyses* (QUOROM)	Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM statement. Lancet 1999;354:1896–900
Metaanálisis de estudios observacionales	*Meta-analyses of Observational Studies in Epidemiology* (MOOSE)	Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008–2012
Revisiones sistemáticas de estudios sobre la precisión diagnóstica	*Quality Assessment of Diagnostic Accuracy Studies* (QUADAS)	Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–536
Estudios de predicción del	*Genetic Risk Prediction Studies*	Janssens AC, Ioannidis JP, van Duijn CM, et al. Strengthening

booksmedicos.org

riesgo genético	(GRIPS)	the reporting of genetic risk prediction studies: the GRIPS statement. Ann Intern Med 2011;154:421–425

Tabla 14-5

Diversos contenidos de una revista médica general	
Ciencia	**Profesión**
Investigación original	Formación médica
Estudios preliminares	Historia
Artículos de revisión	Política pública
Editoriales (de síntesis y opinión)	Revisiones de libros
Cartas al director	Noticias
Hipótesis	Historias y narraciones

«Lectura» de revistas

La capacidad de autocriticar las propias investigaciones es una habilidad esencial para los médicos. Sin embargo, esta aptitud se utiliza de forma selectiva y en diferentes grados, justamente cómo la integridad del interrogatorio y la exploración física, que forman parte del repertorio médico, se utiliza en diferente medida de un paciente al otro. No es necesario leer las revistas de la primera a la última página, igual que nadie leería un periódico desde el principio hasta el final. Se pueden leer por partes, «navegar» según el tiempo del que se disponga y de la validez y relevancia de cada artículo individual.

Las propuestas para que la lectura sea más eficaz varían. Una buena idea puede ser leer, por lo menos, los títulos (análogos a los titulares de un diario) de todos los artículos de un ejemplar para decidir los que más le interesen. Una vez seleccionados, se puede profundizar, ajustando la importancia conforme se lee (fig. 14-2). El resumen es la parte más indicada para empezar, y muchos lectores serios se detienen en este punto. Si las

booksmedicos.org

conclusiones son interesantes, se puede continuar con el apartado de métodos, y luego se puede encontrar información básica en relación a si las conclusiones son verosímiles. Es conveniente leer el apartado de resultados para obtener una descripción más detallada de lo que se ha descubierto. Las figuras clave (como la curva de supervivencia de los resultados principales de un ensayo aleatorizado) pueden comunicar la información básica sobre la potencia total de forma eficiente. Algunos artículos son tan relevantes, en función de las propias necesidades personales, que vale la pena leerlos palabra por palabra, quizá como preparación para reuniones de debates de artículos científicos.

booksmedicos.org

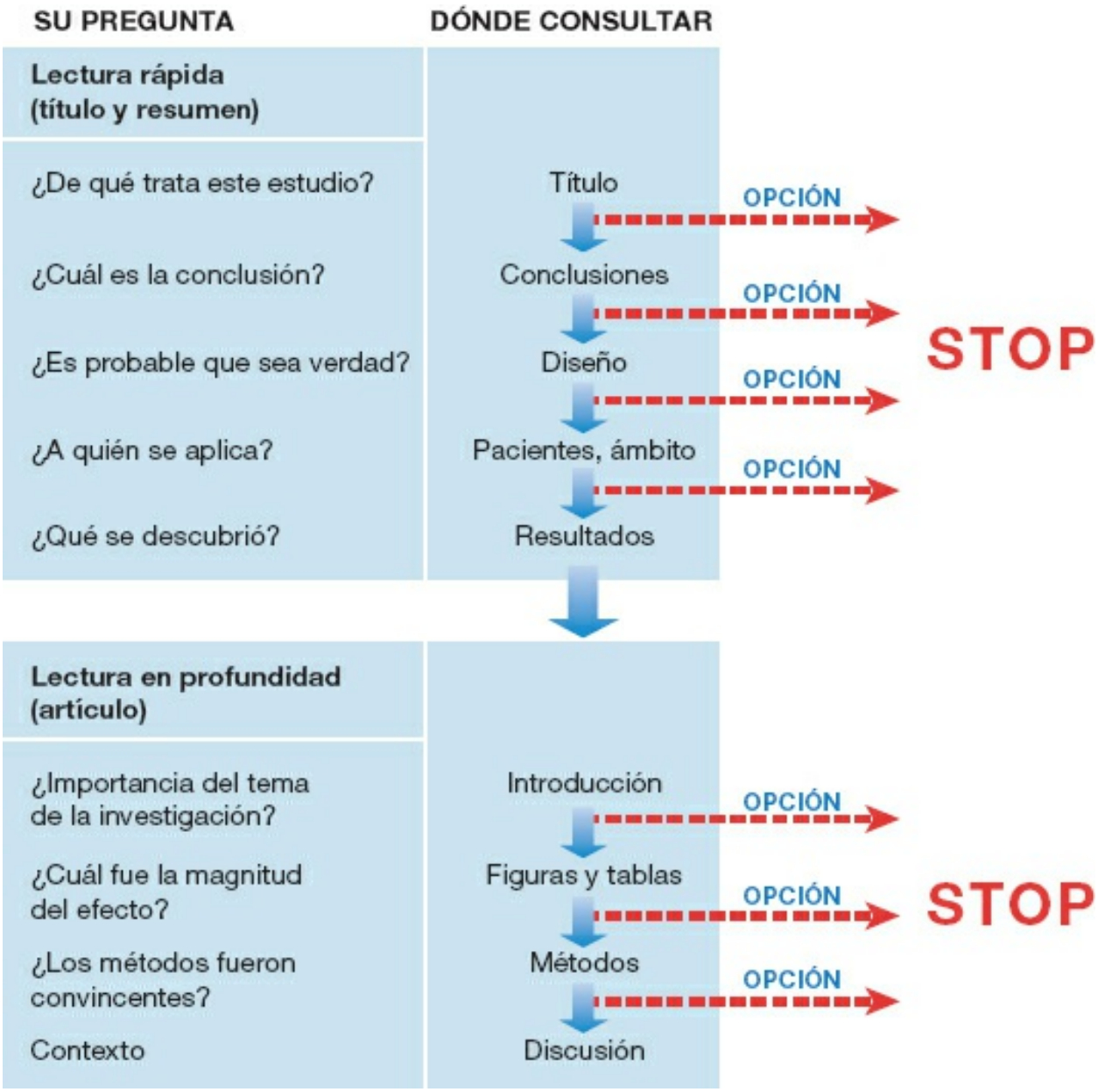


Figura 14-2 ■ **Lectura de un artículo de revista por partes.** El lector puede profundizar en un artículo o dejarlo y leer otro, en función de la validez científica y la importancia clínica que para él tengan los artículos.

Tabla 14-6

Organización de un resumen estructurado	
Título	**Información**
Contexto	¿Impacto del sufrimiento a causa de la enfermedad? ¿Por qué es importante la cuestión de la investigación? ¿Qué es lo que ya se sabe?

booksmedicos.org

Objetivo	Qué desean saber los investigadores
Ámbito	El ámbito en el que se pueden generalizar los resultados, por ejemplo, la sociedad, las prácticas de atención primaria, los centros de referencia, etc.
Participantes	¿Qué tipo de pacientes (en cuanto a la capacidad de generalización)? ¿Cuántos (en cuanto a la potencia/precisión estadística)?
Diseño	¿Cuál es la solidez del estudio? ¿Se aborda con eficacia la cuestión de la investigación?
Intervención (si la hay)	¿La intervención es de vanguardia? ¿Es factible en su ámbito?
Medidas básicas del resultado	¿Son los resultados clínicamente importantes?
Resultados	¿Qué se ha detectado?
Limitaciones	¿Qué aspectos del estudio amenazan la validez de las conclusiones?
Conclusión	¿Los autores consideran que los resultados responden a su pregunta? ¿De manera convincente?

Los **resúmenes estructurados** se organizan según los tipos de información de la que los lectores críticos dependen cuando han de decidir si confían en los resultados o no. La tabla 14-6 muestra los títulos de los resúmenes de una forma estructurada, junto con el tipo de información asociada con estos. (Los resúmenes tradicionales, con encabezados para la Introducción, Métodos, Resultados y Discusión, son una versión abreviada.) Estos títulos facilitan a los lectores encontrar la información que necesitan y también obliga a los autores a incluir esta información en los títulos, parte de la cual se podría haber omitido si el resumen fuera menos estructurado.

Desafortunadamente, muchos médicos se fijan metas para la lectura de artículos que son más altas de lo que pueden lograr, creyendo que deben revisar cada artículo con detalle, lo que comporta mucho tiempo para cada número de la revista. También conlleva mucha ansiedad, remordimientos y espacios de trabajo desordenados. Si esta sensación negativa se asocia a la lectura de las revistas médicas, hay algo que no funciona bien.

booksmedicos.org

GUIAR A LOS PACIENTES PARA BUSCAR INFORMACIÓN SOBRE SALUD

Actualmente, los pacientes buscan información sobre la salud en internet. En consecuencia, los médicos tienen diferentes responsabilidades para instruirles.

Una responsabilidad es indicarles qué sitios web tienen mayor credibilidad. Las búsquedas simples, como la cefalea por migraña o la pérdida de peso, encuentran una amplia variedad de sitios, algunos entre los mejores del mundo, otros entusiastas pero equivocados y otros comerciales y egoístas. Los médicos deben ser capaces de sugerir sitios web especialmente buenos para las preguntas particulares de los pacientes. Existen varios que son subvencionados por el gobierno, escuelas de medicina, organizaciones profesionales y grupos para la defensa del paciente. Los médicos también pueden ayudar a los pacientes a reconocer la mejor información para la salud en internet, guiándose por los criterios formulados por la Medical Library Association (tabla 14-7).

Otra responsabilidad es ayudar a los pacientes a sopesar el valor de la información que encuentran. En este sentido, los médicos tienen mucho que ofrecer debido a su comprensión de la epidemiología clínica, la biología de la enfermedad, las presentaciones clínicas de la enfermedad, la diferencia entre las observaciones aisladas y los patrones consistentes de evidencia, y mucho más. Todo esto es un complemento valioso que puede aportar a la cita, un gran interés sobre un tema clínico específico y la voluntad de invertir mucho tiempo en busca de respuestas.

Tabla 14-7

Criterios que los pacientes pueden usar para evaluar la información sobre la salud en internet

1. Patrocinado
   - ¿Se pueden identificar fácilmente a los patrocinadores del sitio? ¿Se mencionan a los miembros de la junta de asesoramiento y a los consultores?
   - ¿Cuál es la dirección web? (gov = gobierno, edu = institución educativa, org = organización profesional, com = comercial)?

booksmedicos.org

2. Actualización
   • El sitio debe haber sido actualizado recientemente y se debe publicar la fecha de la última revisión
3. Información objetiva
   • La información debe tratar sobre hechos no sobre opiniones, y debe ser verificada de fuentes primarias como artículos profesionales
   • Cuando se expresen opiniones, se debe identificar la fuente (un profesional u organización cualificada)
4. Público
   • El sitio web debe mencionar claramente si la información es para consumidores o profesionales de la salud. (Algunos sitios tienen áreas separadas para los consumidores y los profesionales de la salud)

Modificado de Medical Library Association. A User's Guide to Finding and Evaluating Health Information on the Web. Available at http://mlanet.org/resources/userguide.html. Accessed August 1, 2012.

PONER EN PRÁCTICA LA ADMINISTRACIÓN DE LOS CONOCIMIENTOS

La epidemiología clínica, como se describe en este libro, pretende facilitar la vida de los profesionales sanitarios y satisfacerles. Con una buena base sólida de los principios con los que se juzga la validez y la capacidad de generalización de la información clínica, los médicos pueden detectar con mayor rapidez y precisión si la base científica de las afirmaciones es concluyente. Por ejemplo, los médicos pueden ver si los intervalos de confianza concuerdan con el beneficio clínicamente importante, o perjudican, o que un estudio sobre los efectos de una intervención no incluye la aleatorización ni otros esfuerzos para tratar con la confusión. Además, están mejor preparados para participar en debates con otros médicos de diferente especialidad sobre las decisiones de la atención del paciente y tienen una mejor base para decidir cómo delegar algunos aspectos de sus necesidades de información, y pueden ganar más confianza y experimentar más satisfacción con los aspectos intelectuales de su trabajo.

Más allá de eso, cada médico debe tener un plan para administrar el conocimiento que se adapte a las necesidades y los recursos particulares. Internet debe ser una parte importante del plan porque no hay ningún otro

booksmedicos.org

medio tan completo, actualizado y flexible. Gran parte de la información necesaria para guiar las decisiones sobre la atención del paciente debe estar disponible en el lugar de la atención médica para ponerla en práctica en el paciente en cuestión. No hay ninguna razón para que la información que se utiliza no pueda ser la mejor disponible en el mundo en ese momento, siempre y cuando se tenga acceso a internet.

Un método viable para la administración de los conocimientos debe ser activo. Los médicos deben invertir tiempo en crear un proyecto que cumpla con todas sus necesidades de información. De vez en cuando, deben dedicar un tiempo extra para revisar dicho proyecto, para conocer las nuevas oportunidades, cuando aparecen, y para adquirir las nuevas habilidades que se necesitan. Las evidencias para la medicina clínica nunca habían sido tan sólidas y accesibles como hasta ahora. ¿Por qué no aprovecharlas al máximo?

Preguntas de revisión

Lea los siguientes enunciados y seleccione la mejor respuesta.

14-1. Está terminando su período de formación, pronto empezará a ejercer. Quiere establecer un plan para mantenerse al día de los progresos en su especialidad aunque no haya muchos profesionales sanitarios en su localidad. Los siguientes puntos pueden ser *más* útiles, ¿pero cuál puede ser el más provechoso?

A. Suscribirse a unas cuantas revistas.
B. Comprar las últimas ediciones de libros de texto impresos.
C. Suscribirse a un servicio que revise las publicaciones de su especialidad.
D. Hacer búsquedas en MEDLINE con regularidad.
E. Mantener el contacto con compañeros del programa de formación por correo electrónico y teléfono.

14-2. Puede confiar en las mejores revistas de medicina de su especialidad para:

A. Obtener respuestas a las cuestiones médicas.

booksmedicos.org

B. Asegurarse de que está al día con las publicaciones médicas.
C. Garantizarle que la información que contienen es irreprochable.
D. Exponerle la diversidad de capacidades de su profesión.

14-3. Muchos niños de su consulta tienen otitis media, y usted quiere basar su tratamiento en las mejores pruebas disponibles. ¿Cuál de las siguientes fuentes de información es la menos *fiable* sobre este tema?

A. Una guía de práctica clínica realizada por una sociedad médica prestigiosa.
B. Una revisión sistemática publicada en una revista muy reconocida.
C. La base de datos de revisiones sistemáticas de Cochrane.
D. El artículo de investigación más reciente sobre este tema.

14-4. La búsqueda en MEDLINE es especialmente útil porque:

A. Se encuentran los mejores artículos relacionados con una pregunta clínica.
B. Es una estrategia eficaz para encontrar los mejores artículos.
C. Se pueden consultar informes de episodios poco frecuentes.
D. Se puede estar al día en cuanto a las publicaciones médicas.
E. Es posible familiarizarse con la medicina en general.

14-5. ¿Qué se consigue mediante la revisión por colegas de los manuscritos de investigación antes de su publicación?

A. Excluir los artículos de los autores con conflicto de intereses.
B. Hacer que el artículo publicado sea preciso y fiable.
C. Aliviar la necesidad de los lectores de ser escépticos sobre el estudio.
D. La decisión de los editores sobre si se debe publicar el manuscrito.

14-6. El autor de un artículo que demuestra que la colonoscopia de tamizaje es más efectiva para prevenir el cáncer colorrectal que otras formas de tamizaje podría tener conflictos de intereses si él o ella tuviera lo siguiente, *excepto:*

A. Ingresos médicos por realizar colonoscopias.
B. Inversión en una compañía que realiza colonoscopias.

booksmedicos.org

C. Inversión en productos médicos en general.

D. Publicaciones de artículos que han defendido consistentemente la colonoscopia como la mejor prueba de tamizaje.

E. Rivalidad con otros estudiosos que abogan por otra prueba de tamizaje.

14-7. ¿Cuál de las siguientes es la forma menos útil de buscar las respuestas a las preguntas clínicas en el lugar de atención médica?

A. Suscribirse a varias revistas y tenerlas disponibles en el sitio donde atiende a los pacientes.

B. Consultar guías de práctica clínica en http://www.guidelines.gov

C. La biblioteca Cochrane en internet.

D. Un libro de texto electrónico que se actualiza continuamente.

14-8. ¿Cuál de las siguientes características debe ser la menos tranquilizadora para un paciente sobre la calidad de un sitio web que proporciona información sobre el VIH?

A. El sitio es financiado por una agencia gubernamental y nombra a los miembros de la junta asesora.

B. El sitio proporciona hechos, no opiniones.

C. Se indica la fuente principal de información.

D. El autor es un experto muy reconocido en la materia.

E. La fecha de la última revisión está publicada y es reciente.

14-9. ¿Cuál de las siguientes acciones no forma parte de la clasificación de las recomendaciones clínicas usando el sistema GRADE?

A. Decidir si se usa una prueba diagnóstica.

B. Tener en cuenta el equilibro de los beneficios y los daños.

C. Calificar la calidad de la evidencia científica por separado.

D. Sugerir que tan común y enérgicamente se debe recomendar un tratamiento.

E. Calificar la fuerza de la evidencia y las recomendaciones por separado.

14-10. ¿Cuál de las siguientes características debería incluir un enfoque

booksmedicos.org

integral para administrar el conocimiento en su especialidad?

A. Suscribirse a algunas revistas y explorarlas.
B. Establecer un plan para buscar información en el lugar de la atención médica.
C. Encontrar una publicación que le ayude a mantenerse al día con los nuevos avances en su especialidad.
D. Identificar sitios web que pueda recomendar a sus pacientes.
E. Todos los anteriores.

Encontrará las respuestas en el apéndice A.

BIBLIOGRAFÍA

1. Laine C, Horton, R, DeAngelis CD, et al. Clinical trial registration: looking back and moving ahead. *Lancet* 2007;369:1909– 1911.
2. Fletcher RH, Black B. "Spin" in scientific writing: scientific mischief and legal jeopardy. *Med Law* 2007;26(3):511–525.
3. Shariff SZ, Sontrop JM, Haynes RB, et al. Impact of PubMed search filters on the retrieval of evidence for physicians. *CMAJ* 2012;184:303.
4. Losanoff JE, Sauter ER, Rider KD. Cat scratch disease presenting with abdominal pain and retroperitoneal lymphadenopathy. *J Clin Gastroentrol* 2004;38:300–301.
5. Goodman SN, Berlin J, Fletcher SW, et al. Manuscript quality before and after peer review and editing at Annals of Internal Medicine. *Ann Intern Med* 1994;121:11–21.
6. Lacombe MA, ed. On Being a Doctor. Philadelphia: American College of Physicians; 1995.

booksmedicos.org

Apéndice A

Respuestas a las preguntas para la revisión

CAPÍTULO 1 INTRODUCCIÓN

1-1 **D.** Las muestras pueden dar una impresión errónea de la situación en la población de origen, especialmente si la muestra es pequeña.

1-2 **E.** La generalización de los resultados de un estudio realizado en hombres para la asistencia a mujeres supone que la eficacia de la cirugía para la lumbalgia es la misma para los dos sexos.

1-3 **B.** Los dos grupos de tratamiento no tenían la misma probabilidad de que su dolor se hubiera medido.

1-4 **A.** La diferencia de la recuperación entre los pacientes que se han sometido a una intervención quirúrgica en lugar de recibir asistencia médica puede ser consecuencia de algún otro factor, como la edad, que es diferente entre los dos grupos tratados, y no deberse a la intervención en sí.

1-5 **B.** Estos son sesgos relacionados con la medida del resultado (recuperación fuera del dolor).

1-6 **C.** Las otras enfermedades son un factor de confusión de la relación entre el tratamiento y el resultado; es decir, están relacionadas tanto con el tratamiento como con la recuperación, y puede ser la razón de las diferencias que se observan.

1-7 **C.** La observación de que la histamina es un mediador de la inflamación en la rinitis alérgica primaveral conduce a la hipótesis prometedora de que los antihistamínicos aliviarán los síntomas, pero esta suposición se ha de demostrar en personas que sufren este tipo de afección. Las otras respuestas dan por supuesto mucho más sobre las causas de los síntomas de lo realmente establecido. Por ejemplo, la histamina sólo es uno de los muchos mediadores de la inflamación en rinitis alérgica primaveral.

1-8 **C.** Las muestras, especialmente cuando son pequeñas, pueden

booksmedicos.org

representar erróneamente poblaciones por casualidad.

1-9 **A.** La generalización de los pacientes más jóvenes a los más ancianos es una cuestión de juicio personal en función de la respuesta al tratamiento de los pacientes ancianos igual que los pacientes jóvenes. La validez interna depende de si los resultados son correctos para los pacientes del estudio y no sobre si son correctos para otro tipo de pacientes. Tanto los sesgos como el azar afectan a la validez interna.

1-10 **B.** Si los voluntarios hicieran la misma cantidad de ejercicio que los no voluntarios, esta variable no podría explicar una diferencia en la tasa de ECC. Las diferencias en los grupos (sesgo de selección) o en los métodos utilizados para determinar la ECC (sesgo de medición) podrían explicar este hallazgo.

1-11 **C.** El efecto del fármaco sobre la mortalidad es el asunto más importante que debe determinarse. Es posible que algunos antiarrítmicos puedan aumentar la tasa de muerte súbita (de hecho, esto ha ocurrido). En dichas situaciones, el resultado biológico intermedio de la disminución de las arritmias es un marcador poco fiable del resultado clínico, la muerte súbita.

1-12 **B.** El sesgo de medición es un problema frecuente cuando se pide a los pacientes que recuerden algo que puede estar relacionado con su enfermedad, porque las personas enfermeras podrían recordar con mayor facilidad los acontecimientos precedentes que creen que podrían estar relacionados con su enfermedad.

1-13 **B.** El interrogatorio (medición) sobre el uso de anticonceptivos no fue el mismo en los dos grupos.

1-14 **D.** Un estudio con pocos participantes aumenta la posibilidad de que el azar explique las diferencias entre los grupos.

1-15 **A.** Los diferentes vecindarios a menudo dependen de variables socioeconómicas que están relacionadas con numerosos resultados de salud, lo que puede significar que los dos grupos son diferentes en términos de covariables importantes.

CAPÍTULO 2 FRECUENCIA

2-1 **C.** La población en riesgo es dinámica, las personas entran y la abandonan continuamente, por lo que las tasas en los registros del cáncer deben ser citados como años-persona.

booksmedicos.org

2-2 **A.** Esta es la prevalencia puntual porque los casos se describen en un momento del tiempo durante su vida.

2-3 **A.** No hay seguimiento (sin dimensión temporal) en un estudio de prevalencia.

2-4 **D.** Un muestreo aleatorio o probabilístico es representativo de la población de la que se obtiene la muestra a largo plazo, si la muestra se obtiene de suficientes personas en la población.

2-5 **B.** En un estado continuo, la duración de la enfermedad = prevalencia/incidencia. En este caso es 1/100 dividido por 40/10 000/año = 25 años.

2-6 **D.** Un estudio que comienza con los pacientes existentes que tienen una enfermedad, y mira hacia atrás para determinar su experiencia previa con la enfermedad, no sería un estudio de cohorte, que comienza con pacientes que tienen algo en común (p. ej., una enfermedad de nueva aparición) y los sigue adelante en el tiempo para evaluar episodios de salud subsecuentes.

2-7 **C.** A pesar que el muestreo de cada décimo paciente podría producir una muestra representativa, no sería aleatorio y podría representar erróneamente a la población de la que se obtuvo la muestra.

2-8 **C.** Los niños con otra convulsión formaban parte de la cohorte original de niños con una primera convulsión.

2-9 **E.** Una muestra más grande produciría una estimación de la incidencia que se acerca a la verdadera, al reducir el error aleatorio. No introduciría una diferencia sistemática en la incidencia como lo haría A-D.

2-10 **A.** Es endémico porque está confinado a un área geográfica.

2-11 **E.** Los casos existieron en cualquier momento del período de 3 meses.

2-12 **C.** Las poblaciones dinámicas, como los residentes del estado de Carolina del Norte, están cambiando continuamente por los nacimientos y las muertes, además de la emigración y la inmigración.

2-13 **A.** Los niños en la cohorte tienen en común haber nacido en Carolina del Norte en 2012 y se realizó un seguimiento para detectar escoliosis conforme se desarrolla a lo largo del tiempo.

2-14 **D.** Los estudios de prevalencia no son útiles para las enfermedades de corta duración porque habría pocos casos en ese momento. No miden la incidencia y no pueden proporcionar mucha evidencia que apoye a la causa.

2-15 **E.** Un número como 800 000 sin un denominador asociado no es una

booksmedicos.org

tasa y, por lo tanto, no es ninguna de las tasas de incidencia o prevalencia dadas en A-D.

CAPÍTULO 3 ANORMALIDAD

3-1 **D.** Ordinal

3-2 **B.** Dicotómicas.

3-3 **A.** Intervalo, continuas.

3-4 **E.** Intervalo, discretas.

3-5 **C.** Nominal.

3-6 **C.** Este abordaje, denominado constructo de validez, es una de las formas de establecer la validez de una medición. Considere que las respuestas B y D se relacionan con la fiabilidad, no con la validez.

3-7 **D.** Todas, excepto D, son razones para la variación en las mediciones en un solo paciente, mientras que la D es la variación entre los pacientes.

3-8 **D.** Debido a que las distribuciones clínicas no necesariamente siguen una distribución normal, no se debe definir la anormalidad por si lo hace o no.

3-9 **A.** Las distribuciones naturales pueden o no parecerse a una curva normal.

3-10 **C.** Aunque en última instancia usted y el paciente pueden decidir sobre una prueba de tratamiento con estatinas, esta no es una situación de emergencia. El análisis de colesterol debe repetirse; los valores elevados a menudo son más bajos en pruebas consecutivas. Una prueba con ejercicio y pérdida de peso también puede disminuir el colesterol. A los pacientes que por lo demás están sanos, con un riesgo de enfermedades cardiovasculares a 10 años < 10 %, generalmente no se les prescriben de inmediato fármacos para disminuir el colesterol.

3-11 **B.** La figura muestra una moda (joroba). La mediana y la media son similares entre sí, y ambas están por debajo de 4 000 g.

3-12 **C.** Dos desviaciones estándar comprenden el 95 % de los valores. La distribución no está sesgada. El rango es sensible a los valores extremos.

3-13 **B.** Una desviación estándar abarca aproximadamente 2/3 de los valores alrededor de la media. Observando la figura, 2/3 de los valores alrededor de la media serían de unos 3 000-4 000 g.

booksmedicos.org

3-14 **D.** La parte A de la figura 3-13 muestra una dispersión casi uniforme de las mediciones por encima y por debajo del valor real, lo que sugiere una posibilidad de variación. El efecto de los diferentes observadores que realizan las mediciones (variabilidad interobservador) también podría estar en juego.

3-15 **G.** La parte B muestra las mediciones sesgadas a la derecha del valor verdadero. En otras palabras, el personal hospitalario tendió a sobrestimar los resultados de la monitorización electrónica, demostrando que las mediciones fueron sesgadas. El azar y la variabilidad interobservador también pueden participar porque no todas las mediciones son iguales.

3-16 **G.** La parte C muestra resultados similares a la parte B, excepto que el sesgo va en otra dirección, es decir, el hospital hospitalario tendió a subestimar los resultados de la monitorización electrónica. En las partes B y C, las mediciones del personal hospitalario tendieron a «normalizar» las medidas anormales del monitor electrónico.

CAPÍTULO 4 RIESGO: PRINCIPIOS BÁSICOS

4-1 **A.** Cuando un factor de riesgo es común en una población, es importante comparar su prevalencia en las personas con y sin la enfermedad. No hay comparación en el ejemplo. El cáncer de pulmón es común, y el tabaquismo es un factor de riesgo importante, no débil, para el cáncer. El hecho de que no haya otros factores de riesgo para el cáncer de pulmón no es relevante.

4-2 **B.** Los factores de riesgo se buscan cuando una nueva enfermedad aparece, como en el caso del VIH.

4-3 **D.** Los modelos de predicción de riesgo se utilizan en todas las situaciones.

4-4 **C.** Debido a que muchas más mujeres se asignaron al estrato de bajo riesgo, es probable que un mayor número de casos provengan de ese grupo. Los modelos de predicción de riesgo dan probabilidades, pero no predicen qué personas dentro de un grupo desarrollarán la enfermedad. La figura no da información sobre las asignaciones incorrectas a los estratos.

4-5 **C.** Es improbable que el paciente desarrolle cáncer colorrectal en los siguientes 5 años porque es un miembro del grupo de baja probabilidad,

booksmedicos.org

pero podría ser uno del 2 % que desarrollan la enfermedad. Incluso para un cáncer común existen subgrupos de personas con baja probabilidad de desarrollarlo. Por ejemplo, pocas personas jóvenes desarrollan cáncer colorrectal.

4-6 **B.** Es más probable que los modelos de riesgo bien construidos estén altamente calibrados (capaces de predecir el porcentaje de un grupo que desarrolla la enfermedad), en vez de tener una buena discriminación (capaces de predecir qué individuos del grupo desarrollarán la enfermedad).

4-7 **A.** Los marcadores ayudan a identificar los grupos con mayor riesgo de enfermedad, pero ya que no son causas de la enfermedad, eliminando el marcador no se previene la enfermedad. Los marcadores a menudo se confunden con la causa de la enfermedad (v. caps. 1 y 5).

4-8 **B.** Los síntomas, los hallazgos de la exploración física y las pruebas de laboratorio son generalmente más importantes que los factores de riesgo cuando se diagnostica una enfermedad.

4-9 **D.** La superposición de los dos grupos demuestra que el modelo de riesgo no discrimina bien entre las mujeres que desarrollan o no cáncer de mama a los 5 años. La figura no da ninguna información directa sobre la calibración y la estratificación. Pocas mujeres que desarrollaron cáncer de mama tenían alto riesgo.

4-10 **C.** Una buena calibración no afecta a la discriminación. Sin embargo, los modelos de riesgo pueden discriminar pobremente si están altamente calibrados, como muestra el ejemplo en el texto.

CAPÍTULO 5 RIESGO: EXPOSICIÓN A LA ENFERMEDAD

5-1 **C.** Los estudios de cohorte retrospectivos no permiten a los investigadores el lujo de decidir qué datos recopilar, sólo pueden elegir entre los datos disponibles recogidos en un momento antes de iniciar el estudio. Por otro lado, estos datos, que a menudo son recogidos para fines clínicos, pueden no haberse recopilado sistemáticamente y de la misma manera. Las respuestas A, B y D son correctas tanto para los estudios de cohorte prospectivos y retrospectivos.

5-2 **B.** El riesgo relativo es el cociente de la incidencia de un resultado (enfermedad) de los expuestos entre los no expuestos. Por lo tanto, el

booksmedicos.org

riesgo relativo de enfermedad cerebrovascular (ECV) en los fumadores a sus 40 años, en comparación con los no fumadores, es 29,7/7,4 = 4,0.

5-3 **D.** El riesgo atribuible (RA) es el riesgo de una enfermedad atribuible al factor de riesgo y se calcula como la diferencia del riesgo absoluto (incidencia) de las personas expuestas menos el riesgo absoluto en las personas no expuestas. Por lo tanto, el riesgo atribuible de ECV en los fumadores a sus 60 años, en comparación con los no fumadores, es de 110,4 – 80,2 = 30,2.

5-4 **C.** El riesgo relativo (RR) no da información sobre la incidencia, mientras que el riesgo atribuible lo hace, por lo que la información en C es incorrecta. Todas las demás respuestas son correctas. Para calcular el riesgo atribuible a la población se debe conocer la prevalencia del factor de riesgo en la población, que no se especifica en la pregunta. La incidencia de ECV entre los fumadores fue mayor en las personas de 60 años que a los 40 años. El RR de ECV entre los fumadores a sus 40 años fue 4,0 en comparación con 1,4 a los 60 años. Los RR más fuertes son una mejor evidencia para una relación causal que los más débiles. En el análisis de este estudio, la edad podría tratarse como una variable de confusión y ser controlada en consecuencia o, como en la información que se presenta, ser un efecto modificador que muestra que el efecto del tabaquismo varía con la edad.

5-5 **A.** El riesgo absoluto es otro término para incidencia. La trombosis venosa profunda (TVP) es un episodio infrecuente para la mayoría de las mujeres, y la incidencia en este estudio fue 0,8/10 000/ año entre las mujeres que no tenían la mutación o no tomaron anticonceptivos orales (ACO).

5-6 **C.** Una buena forma para organizar su pensamiento para las preguntas 5-6 a 5-10 es crear una tabla de 2 x 2 sobre la incidencia de TVP según el uso de ACO (presente/ausente) y el factor V de Leiden (presente/ausente).

booksmedicos.org

Incidencia (por 10 000/año) de TVP de acuerdo con el uso de ACO y la mutación del factor V de Leiden

		Uso de ACO	
		Presente	**Ausente**
Factor V de Leiden	*Presente*	28,5	5,7
	Ausente	3,0	0,8

La tabla muestra que entre las mujeres que no son portadoras del factor V de Leiden, el riesgo atribuible de TVP por tomar ACO, en comparación con aquellas que no toman ACO, es de 3,0/10 000 – 0,8/10 000 = 2,2/10 000.

5-7 **E.** El riesgo atribuible de TVP entre las usuarias de ACO, que también son portadoras del factor V de Leiden, es importante si se compara con la mayoría de las mujeres que usan ACO, en 28,5/10 000 – 3,0/10 000 = 25,5/10 000.

5-8 **B.** El riesgo atribuible a la población = riesgo atribuible (25,5/10 000 años/mujeres) x prevalencia del factor V de Leiden (0,05) = 1,3/10 000 años/ mujeres.

5-9 **C.** El riesgo relativo se calcula dividiendo la incidencia de TVP entre las mujeres portadoras de la mutación del factor V de Leiden que usan ACO entre la incidencia de TVP en las mujeres que toman ACO pero que no son portadoras de la mutación (28,5/10 000/año entre 3,0/10 000 = 9,5).

5-10 **A.** Entre las mujeres sin la mutación, el riesgo relativo de TVP y el uso de ACO se calcula dividiendo la incidencia de las usuarias de ACO entre la incidencia de las que no usan ACO (3,0/10 000/año entre 0,8/10 000/año = 3,8).

5-11 **C.** Un ejemplo de un factor de riesgo con un riesgo relativo sustancial es que una mujer con una mutación heterocigota del factor V Leiden tome ACO (la prescripción de ACO a este tipo de mujeres aumenta la posibilidad que tienen de sufrir TVP con un riesgo relativo de 5,0), pero un riesgo absoluto relativamente pequeño: 28,5 de 10 000 mujeres que tienen la mutación y utilizan este tipo de anticonceptivos sufrirían TVP el siguiente año. Aún más relevante para estas pacientes es que su riesgo absoluto aumentaría de unos 6 a 28 por 10 000 durante el próximo año. Es importante que una mujer portadora y que desee tomar ACO entienda que su riesgo aumenta y que sepa que lo que realmente incrementa es el

booksmedicos.org

riesgo en términos absolutos. Un médico prudente también querría asegurarse de que la paciente no presenta otras indicaciones de aumento de riesgo de trombosis, como edad, tabaquismo, antecedentes familiares o personales o problemas de coagulación. Este tipo de decisión requiere un juicio minucioso tanto del paciente como del médico. Sin embargo, la utilización del riesgo absoluto o atribuible aclarará el riesgo al paciente mejor que el uso del riesgo relativo cuando se comenten las consecuencias clínicas.

5-12 **B.** El grado de la enfermedad puede haber confundido los resultados en este estudio, por lo que es más probable que los pacientes más enfermos tomen aspirina. Una forma de examinar esta posibilidad y ajustar la confusión, si está presente, es estratificar los usuarios y no usuarios en grupos con indicaciones similares para el uso de aspirina y comparar las tasas de mortalidad en los subgrupos.

CAPÍTULO 6 RIESGO: DE LA ENFERMEDAD A LA EXPOSICIÓN

6-1 **B.** Si la exposición a los anticonceptivos orales se registró justo antes del infarto de miocardio, no podría haber sido la causa de este. Todas las demás opciones podrían haber aumentado artificialmente la exposición en los casos, haciendo que la razón de momios sea falsamente elevada.

6-2 **E.** Incluso un estudio de casos y controles ejemplar como este no podría asegurar que se ha identificado una causa porque siempre es posible la presencia de factores de confusión no medidos.

6-3 **E.** Los estudios de casos y controles producen sólo razón de momios, que se pueden utilizar para estimar el riesgo relativo.

6-4 **C.** Las curvas epidémicas describen el aumento y la disminución del número de casos durante el brote.

6-5 **A.** Siempre se puede obtener un riesgo relativo crudo de un estudio de cohorte. Sin embargo, si los datos de la cohorte no contienen todas las variables que debieron ser controladas, es más eficiente el análisis de casos y controles en el estudio de cohorte para incluir datos adicionales, pues los datos necesitan ser recopilados solamente para los casos y controles, no para toda la cohorte.

6-6 **D.** Sería mejor obtener una muestra de los casos y los controles de una

booksmedicos.org

cohorte en lugar de una población dinámica, sobre todo si la exposición o la enfermedad está cambiando rápidamente a lo largo del tiempo y si los controles no están pareados con la fecha en que inició la enfermedad en los casos.

6-7 **E.** Los grupos de control múltiples (que no deben ser confundidos con múltiples controles por caso) son una forma de examinar si los resultados son «sensibles» a los tipos de controles que se eligieron, es decir, si los resultados son sustancialmente diferentes usando diferentes grupos de control, poniendo en duda los resultados.

6-8 **C.** Los estudios de casos y controles no proporcionan información sobre la incidencia (aunque si están anidados en una cohorte se podría obtener con el análisis de cohorte de los mismos datos).

6-9 **B.** El apareamiento se utiliza para controlar variables que podrían estar fuertemente relacionadas con la exposición o la enfermedad para asegurarse que al menos los casos y los controles no son diferentes en cuanto a esas variables.

6-10 **C.** La razón de momios cruda se obtiene creando una tabla de 2 × 2 que relaciona el número de casos frente a los controles con el número de personas expuestas contra no expuestas, para después dividir sus productos cruzados. En este caso, la razón de momios es 60 × 60 dividido por 40 × 40 = 2,25.

6-11 **E.** Los estudios de casos y controles no pueden estudiar resultados múltiples porque comienzan con la presencia o ausencia de sólo una enfermedad, no pueden reportar la incidencia, y los casos deben ser incidentes (nueva aparición), no prevalentes.

6-12 **D.** Las razones de momios basadas en los casos prevalentes pueden proporcionar una medida aproximada de la asociación, pero no una comparación del riesgo, que trata sobre los casos incidentes (nueva aparición).

6-13 **D.** Durante las fases tempranas de un brote, generalmente se conoce el microorganismo o toxina ofensiva, pero incluso de no ser así, la pregunta más urgente es cómo se disemina la enfermedad. Esta información se puede utilizar para detener el brote e identificar la fuente.

6-14 **C.** Si un estudio de casos y controles se basa en todos o en una muestra aleatoria de los casos y los controles de una población o una cohorte, los casos y los controles deben ser similares entre sí en cuanto al resto de las características aparte de la exposición.

booksmedicos.org

6-15 **D.** Los estudios de casos y controles no proporcionan información sobre la incidencia.

6-16 **E.** La razón de momios se aproxima al riesgo relativo cuando la enfermedad es rara; la regla de oro es < 1/100.

CAPÍTULO 7 PRONÓSTICO

7-1 **A.** El tiempo cero es el inicio de la enfermedad (en este caso, el esófago de Barrett), no el momento de los efectos resultantes (en este caso, cáncer de esófago).

7-2 **C.** Las tasas diferentes de abandono podrían sesgar los resultados sólo si los que lo hubieran abandonado tuvieran un pronóstico sistemáticamente diferente de aquellos que permanecieron en el estudio.

7-3 **E.** Las respuestas A-D son todas posibles razones para un sesgo de medición, mientras que una diferencia verdadera en las tasas de incontinencia entre los pacientes tratados con cirugía contra el tratamiento médico no sería un error sistemático.

7-4 **E.** Las respuestas A-D son características de las reglas de predicción clínica.

7-5 **C.** Las series de casos son una estrategia de investigación híbrida, no se describe el curso clínico de una cohorte, ni el riesgo relativo en el método de casos y controles ni se estudia una muestra representativa de los casos prevalentes.

7-6 **C.** Los pacientes podrían ser censurados por cualquier motivo que los excluya del estudio o que haga que no estén dentro del estudio durante tanto tiempo como 3 años. Si el resultado es la supervivencia, no importaría si tuvieran otra enfermedad mortal si aún no han muerto por este motivo.

7-7 **A.** El pronóstico trata sobre los resultados de la enfermedad a lo largo del tiempo y los estudios de prevalencia no miden episodios a lo largo del tiempo.

7-8 **B.** Las curvas de supervivencia estiman la supervivencia en una cohorte, después de tomar en consideración los pacientes censurados y no describen directamente la proporción de la cohorte original que sobrevivió.

7-9 **E.** Todas las respuestas podrían ser correctas según la pregunta de investigación. Por ejemplo, sería útil conocer el pronóstico tanto en los

booksmedicos.org

pacientes de atención primaria y también entre los referidos a los especialistas.

7-10 **B.** Debido a que los pacientes deben cumplir criterios estrictos para ser incluidos en el ensayo, el curso clínico de los asignados al cuidado habitual no sería representativo de los pacientes con la enfermedad, como ocurre en la población general o en un ámbito clínico definido. Por lo tanto, los resultados no podrían generalizarse a ningún grupo natural de pacientes con esclerosis múltiple.

7-11 **B.** Incluso si se construyeran reglas de predicción clínica utilizando los mejores métodos disponibles, la prueba más fuerte para definir su capacidad de predicción es que lo hayan demostrado en otros pacientes diferentes a los que se utilizaron para crear la regla.

7-12 **D.** El riesgo relativo se calcula a partir de la información en una curva de supervivencia y es similar, pero no idéntico, al riesgo relativo.

7-13 **C.** Los episodios de resultado en los análisis del tiempo hasta el acontecimiento son ambos/o episodios que ocurren sólo una vez.

CAPÍTULO 8 DIAGNÓSTICO

Primero, determine los números para cada una de las cuatro celdas de la figura 8-2: a = 49; b = 79, c = 95 – 49 (o 46); d = 152 – 79 (o 73). Agregue los números para determinar los totales de las columnas y las filas.

8-1 **C. 52 %** Sensibilidad = 49/95 = 52 %.

8-2 **B. 48 %** Especificidad = 73/152 = 48 %.

8-3 **A. 38 %** Valor predictivo positivo = 49/128 = 38 %.

8-4 **D. 61 %** Valor predictivo negativo = 73/119 = 61 %.

8-5 **A. 38 %** Prevalencia de la sinusitis en esta consulta = 95/247 = 38 %.

8-6 **A. 38 %**

1. Calcule el LR de una prueba positiva (el paciente tiene dolor facial).

$$LR+ = \frac{49/(49 + 46)}{79/(79 + 73)} = 1{,}0$$

2. Convierta el momio preprueba en momios *(odds)* preprueba.

booksmedicos.org

$$\text{Momio preprueba} = \text{prevalencia}/1 - \text{prevalencia}$$
$$= 0{,}38/(1 - 0{,}38)$$
$$= 0{,}61$$

3. Calcule los momios posprueba multiplicando los momios preprueba por LR+

$$\text{Momio posprueba} = 0{,}61 \times 1{,}0$$
$$= 0{,}61$$

4. Convierta los momios posprueba a la probabilidad posprueba

$$\text{Momio posprueba} = \text{momio posprueba}/(1 + \text{momio posprueba})$$
$$= 0{,}61/(1 + 0{,}61)$$
$$= 0{,}38\,\% \text{ o } 38\,\%$$

(Otro enfoque más simple es el momio posprueba = valor predictivo positivo)

$$+\text{PV} = 49/128$$
$$= 38\,\%$$

8-7 **D. ~75 %** El LR para «alta probabilidad» de sinusitis fue 4,7 y la probabilidad preprueba (prevalencia) fue del 38 %. Puede utilizar cualquiera de los tres métodos para determinar la probabilidad posprueba de la sinusitis entre los pacientes con «alta probabilidad»: (1) el enfoque matemático descrito en la figura 8-8 (o en internet), (2) usando un nomograma o (3) usando el método de la «regla de oro» descrito en la tabla 8-3.

1. Enfoque matemático:

$$\text{Momio preprueba} = 0{,}38/(1{,}38) = 0{,}61$$
$$\text{LR+} \times \text{momio preprueba} = 4{,}7 \times 0{,}61 = 2{,}9$$
$$\text{Momio posprueba} = 2{,}9/(1 + 2{,}9\,\%)$$
$$= 0{,}74\,\% \text{ o } 74\,\%$$

2. Nomograma

Coloque una regla en el 38 % para la prevalencia y en 4,7 para el LR; cruza el número posprueba en un 75 %.

3. «Regla de oro» de cabecera

booksmedicos.org

Un LR de 4,7 (cercano a 5) aumenta la probabilidad de sinusitis unos 30 puntos porcentuales, de 38-68 %.

8-8 **C. ~45 %.**

8-9 **B. ~20 %.**

8-10 **C.** Conforme la prevalencia disminuye, el valor predictivo de la prueba también (v. fig. 8-7). Es más probable que los médicos traten a un paciente con una probabilidad de sinusitis del 75 % a que renuncien al tratamiento ante una probabilidad del 20 %. (En este caso pueden ser necesarios otros estudios, como radiografías de senos paranasales). La probabilidad posprueba de «probabilidad intermedia» de sinusitis es del 45 %, casi como lanzar una moneda al aire.

8-11 **D.** Al solicitar que dos pruebas independientes sean anormales, es decir, usarlas en serie, aumenta la especificidad y el valor predictivo positivo. (V. ejemplo en la pág. 126-7 y la tabla 8-4). Sin embargo, este abordaje disminuye la sensibilidad; por lo tanto, B es incorrecta. El uso de prueba en paralelo aumenta la sensibilidad (por lo que A es incorrecta), pero en general disminuye el valor predictivo positivo (haciendo que C sea incorrecta). Trabaje con la figura 8-2 y 8-3 para convencerse.

8-12 **A.** El requerimiento más importante cuando se utilizan múltiples pruebas es que cada una aporte información independiente que no se haya hecho evidente en una prueba anterior. Cuando se utilizan pruebas en serie, lo más eficiente es realizar primero la prueba con la mayor especificidad y requiere que menos pacientes hagan ambas pruebas. En el caso de pruebas en paralelo, lo más eficiente es realizar la prueba con la mayor sensibilidad.

CAPÍTULO 9 TRATAMIENTO

9-1 **A.** Se hizo el esfuerzo necesario para que este ensayo se acercara al máximo a la vida real, comparando fármacos de uso común, manteniendo criterios de elegibilidad amplios, no cegando a los participantes, permitiendo que la atención procediera conforme a lo habitual y confiando en un resultado centrado en el paciente en vez de una medición de laboratorio. Aunque es un ensayo de eficiencia, se describe mejor como práctico y ciertamente no es grande.

9-2 **D.** El análisis por intención de tratar, en el que se consideran los resultados de acuerdo con el grupo de tratamiento al que fueron

booksmedicos.org

asignados los pacientes, prueba los efectos de ofrecer el tratamiento, sin importar si realmente los pacientes lo tomaron. Por lo tanto, se trata de la efectividad en circunstancias ordinarias y la medida del efecto que, al igual que la atención habitual, se ve afectada por los abandonos que reducen el efecto del tratamiento observado en comparación con lo que habría sido si todos hubieran tomado el tratamiento al que fueron asignados.

9-3 **A.** Todas las características en el momento de la aleatorización, como la gravedad de la enfermedad, se asignan al azar. Las características que surgen después de la aleatorización, como la retención, la retención, la respuesta al tratamiento y el cumplimiento, no se asignan al azar.

9-4 **D.** La mayor ventaja que tienen los ensayos aleatorizados sobre los estudios observacionales es prevenir la confusión. La aleatorización crea grupos de comparación que tendrían las mismas tasas de resultados, de promedio, si no fuera por los efectos de la intervención.

9-5 **E.** El estudio tiene criterios de inclusión y exclusión amplios. Esto aumentaría el grado en que los participantes del ensayo fueron similares entre sí, por lo que es más fácil detectar diferencias si existieran, pero a expensas de la generalización, es decir, la capacidad de extrapolar los resultados del estudio a la atención habitual de los pacientes.

9-6 **A.** El análisis por intención de tratar describe los efectos de ofrecer los tratamientos, no sólo de tomarlos. Para describir los efectos de recibir realmente la intervención, se deberían tratar los datos como si fueran de un estudio de cohorte y usar una variedad de métodos para controlar la confusión.

9-7 **B.** La aleatorización estratificada es un método para controlar la confusión, especialmente útil cuando una característica se relaciona de forma sólida con el resultado y también cuando el estudio es demasiado pequeño como para pensar que la aleatorización probablemente no haya formado grupos con un pronóstico similar.

9-8 **C.** En los análisis aclaratorios, los resultados se atribuyen al tratamiento que los pacientes realmente reciben, no al grupo de tratamiento al que fueron aleatorizados, que es un análisis por intención de tratar.

9-9 **B.** Los efectos adversos del fármaco, tanto síntomas como signos, podrían alertar a los pacientes y a los médicos sobre quién está tomando el fármaco activo, pero no afectaría a la asignación aleatoria, que se realizó antes de iniciar el fármaco.

booksmedicos.org

9-10 **C.** Para que un ensayo aleatorizado controlado sea ético, no debe existir evidencia concluyente de que uno de los tratamientos experimentales es mejor o peor que el otro, es decir, la comunidad científica debe estar en un estado de «equilibrio» en cuanto a ese tema. Puede haber opiniones sobre la superioridad de uno, pero no un consenso. Es posible que haya cierta evidencia en relación a la comparación, pero no debería ser concluyente.

9-11 **D.** Debido a que el nuevo fármaco tiene ventajas sobre el anterior, pero se desconoce la efectividad comparativa, el ensayo aleatorizado apropiado para compararlos sería un ensayo de no inferioridad para establecer si el nuevo fármaco no es menos eficaz.

9-12 **A.** Haciendo que el resultado principal de un ensayo sea una composición de resultados clínicamente importantes y relacionados, aumenta el número de efectos de resultado y, por lo tanto, la capacidad del ensayo para detectar un efecto de estar presente. La desventaja es que la intervención puede afectar de forma diferente a los resultados que lo componen, y confiar sólo en el resultado compuesto podría enmascarar este efecto.

9-13 **C.** Tanto la mala suerte en la aleatorización como revelar el encubrimiento de la asignación (por métodos malos o trampas) se manifestarían como diferencias en las características basales de los pacientes en un ensayo. Se esperan diferencias pequeñas y el reto es decidir qué tan grandes deben ser para que se consideren preocupantes.

9-14 **C.** Los ensayos farmacológicos habituales reportados en la literatura clínica como ensayos de «fase III» están destinados a establecer la eficacia o efectividad. Son necesarios otros estudios, con vigilancia posterior a la comercialización, para detectar efectos adversos poco habituales. Las respuestas A y D son acerca del propósito de los ensayos de fase I y fase II.

9-15 **B.** La prevención de la confusión es la ventaja principal de los ensayos aleatorizados controlados, motivo por el cual son apreciados a pesar de ser éticamente complicados, más lentos y más costosos. Dependiendo del diseño del ensayo pueden parecerse a la atención habitual.

CAPÍTULO 10 PREVENCIÓN

10- **A.** 33 %. La reducción del riesgo relativo en la mortalidad por cáncer

booksmedicos.org

1 colorrectal es la reducción del riesgo absoluto dividida entre la tasa de mortalidad por cáncer en el grupo control (v. cap. 9).

Tasa de mortalidad por cáncer colorrectal en el grupo de tamizaje = 82/15 570 = 0,0053

Tasa de mortalidad por cáncer colorrectal en el grupo control = 121/15 394 = 0,0079

Reducción del riesgo absoluto = 0,0079 – 0,0053 = 0,0026

Reducción del riesgo absoluto de la mortalidad por cáncer colorrectal a causa del tamizaje = 0,0026/0,0079 = 0,33 o una reducción del 33 %. Un enfoque alternativo calcula el complemento del riesgo relativo. En el ejemplo, el riesgo relativo de la mortalidad colorrectal en el grupo de tamizaje en comparación con el control fue 0,0053/0,0079 = 0,67. El complemento de 0,67 = 1,00 – 0,67 = 0,33 % o 33 %.

10-2 **C.** 385. El número que es necesario cribar es el recíproco de la reducción del riesgo absoluto, 1/0,0026, o 385.

10-3 **C.** Si se encontró que el 30 % del grupo de tamizaje tuvo pólipos, al menos 4 671 (0,3 x 15 570) tuvieron una prueba falsa positiva para cáncer de colon. Si la sensibilidad de la prueba fue del 90 % y el número de carcinomas fue del 82 %, unas 74 personas (0,9 × 82) tuvieron un resultado verdadero positivo. Por lo tanto, el valor predictivo positivo de la prueba sería de un 74/(74 + 4 671) o 1,6 %. (Utilizando número exactos, los autores calcularon un valor predictivo positivo de 2,2 %.) El valor predictivo negativo no fue calculado, pero sería elevado porque la incidencia de cáncer en 13 años fue baja (323/15 570 o 21 por mil) y un 90 % de las pruebas fueron negativas.

10-4 **B.** El tiempo de anticipación (el período entre la detección de la enfermedad en el tamizaje y cuando normalmente sería diagnosticada porque el paciente busca atención médica debido a los síntomas) puede estar asociado con lo que parece ser una mejoría en la supervivencia. El hecho de que la mortalidad no mejoró tras el tamizaje en este ensayo aleatorizado plantea la posibilidad de que el sesgo de anticipación diagnóstica sea responsable del resultado. Otra causa posible del hallazgo es el sobrediagnóstico.

10-5 **A.** Véase la respuesta para **10-4**.

booksmedicos.org

10-6 **C.** Es probable el sobrediagnóstico, la detección de lesiones que no habrían causado síntomas clínicos o morbilidad, porque después de 20 años el número de carcinomas en el grupo control fue menor que en el grupo de tamizaje. Podrían presentarse un mayor número de carcinomas en el grupo de tamizaje si hubiera más fumadores en este grupo, pero la aleatorización debería haber hecho que esa posibilidad fuera improbable.

10-7 **D.** Se deben considerar varias preguntas importantes cuando se evalúa una nueva vacuna, incluyendo: *i)* ¿La vacuna ha demostrado ser eficaz (funcionó bien en circunstancias ideales cuando todos en el grupo de intervención recibieron la vacuna y ninguno la recibió en un grupo control comparable) y fue efectiva (en circunstancias del día a día, donde algunas personas del grupo de intervención no recibieron la vacuna)?, *ii)* ¿La vacuna es segura?, *iii)* ¿El padecimiento contra el que protege la vacuna produce una carga de sufrimiento lo suficientemente importante como para considerar una medida preventiva? y *iv)* ¿Es costo-efectivo? El costo de la vacuna es sólo un componente del análisis de costo-efectividad.

10-8 **C.** La prevalencia de la enfermedad es menor en personas presumiblemente sanas (tamizaje) que en pacientes sintomáticos (diagnóstico). Como resultado, un valor predictivo positivo sería más bajo en el tamizaje. Debido a que el tamizaje está orientado a detectar la enfermedad temprana, la sensibilidad de la mayoría de las pruebas es menor en el tamizaje que en los pacientes con una enfermedad más avanzada. El sobrediagnóstico es menos probable en situaciones diagnósticas, cuando los pacientes sintomáticos tienen una enfermedad en una etapa más avanzada.

10-9 **C.** Los voluntarios para la atención preventiva adoptan más los consejos sobre la atención médica y en general tienen mejores resultados de salud que los que rechazan la atención preventiva. Este efecto es tan fuerte que se observa incluso cuando los voluntarios toman fármacos placebo.

10-10 **B.** El método de la incidencia es particularmente útil cuando se calcula la sensibilidad para las pruebas de tamizaje, pues considera la posibilidad de sobrediagnóstico. La prueba de referencia para las pruebas de tamizaje casi siempre comprenden un intervalo de seguimiento. La primera ronda de tamizaje detecta los casos

booksmedicos.org

prevalentes además de los incidentales, inflando el número comparado con las rondas posteriores de tamizaje.

10-11 **E.** El costo-efectividad se debe estimar de una forma que capture todos los costos de la actividad preventiva y todos los que están asociados con el diagnóstico y el tratamiento, sea donde sea que se produzcan, para determinar los costos de la actividad preventiva desde una perspectiva social. Todos los gastos que se producen cuando no se lleva a cabo la prevención se restan de todos los gastos cuando se realiza la prevención, para estimar el costo de un determinado efecto sobre la salud.

CAPÍTULO 11 AZAR

11-1 **C.** Aunque el análisis por subgrupos tiene el riesgo de obtener conclusiones falsas positivas y falsas negativas, proporciona información que puede ayudar a los médicos, siempre y cuando se tengan en cuenta sus limitaciones.

11-2 **C.** El valor de *P* describe el riesgo de una conclusión falsa positiva y no tiene nada que ver directamente con el sesgo y la potencia estadística con la generalización o la importancia clínica del hallazgo.

11-3 **E.** El valor de *P* no es lo suficientemente pequeño para establecer que el efecto del tratamiento existe y no proporciona ninguna información sobre si un efecto clínicamente importante podría haberse pasado por alto debido a una potencia estadística inadecuada.

11-4 **D.** El resultado fue estadísticamente significativo si se quisiera pensar en el papel del azar de esa manera, pues excluye el riesgo relativo de 1,0. La respuesta D describe la información con la que contribuye el intervalo de confianza.

11-5 **B.** Llamar a $P \leq 0{,}05$ como «estadísticamente significativa» es una convención habitual, pero por lo demás no tiene un significado clínico o matemático en particular.

11-6 **A.** Los modelos dependen de las conjeturas sobre los datos, no se realizan de una forma estándar y están destinados a complementar los análisis estratificados, no a reemplazarlos. Aunque se podrían usar en ensayos aleatorizados grandes, no son particularmente útiles en esa situación porque la aleatorización de un gran número de pacientes ya ha hecho que la confusión sea muy improbable.

booksmedicos.org

11-7 **C.** De las 12 000 personas en el ensayo, alrededor de 6 000 estarían en el grupo de prevención farmacológica del ensayo. Aplicando la regla de oro mencionada en este capítulo (resuelto para el evaluador del episodio, no para el tamaño de la muestra), 6 000 dividido entre 3 o una tasa de episodios de 1/2 000 podría ser detectado con 6 000 personas en observación.

11-8 **E.** La potencia estadística depende de los efectos conjuntos de todos los factores mencionados de A-D. Puede variar un poco con la prueba estadística utilizada para calcularlo, pero esto no es el determinante principal del tamaño de la muestra.

11-9 **B.** El razonamiento bayesiano es sobre cómo afecta la nueva información a la creencia previa y no tiene nada que ver con la estadística inferencial o el fundamento ético de los ensayos aleatorizados.

11-10 **D.** Los resultados son consistentes con una tasa de mortalidad del 1 % hasta el 46 % mayor, con una mejor estimación del 22 % mayor y excluye un cociente de daño de 1,0 (sin efecto), por lo que es «estadísticamente significativo». Los intervalos de confianza proporcionan más información que un valor de P para los mismos datos porque incluyen la estimación puntual y el rango de valores que probablemente incluyan el efecto verdadero.

CAPÍTULO 12 CAUSA

12-1 **B.** Sería poco ético asignar al azar una actividad potencialmente dañina, y es probable que las personas no aceptaran que se le restringiera durante mucho tiempo el uso de sus teléfonos móviles.

12-2 **E.** Todas las respuestas, que reflejan los criterios de Bradford Hill, podrían ser útiles para decidir si los teléfonos móviles causan cáncer cerebral, pero E, la falta de especificidad, es más débil que el resto. (Considere cuántas enfermedades causan el tabaquismo.)

12-3 **C.** La figura 12-1 muestra cómo varios factores de riesgo importantes actúan en conjunto para causar una enfermedad cardíaca coronaria. Casi todas las enfermedades están causadas por los efectos conjuntos de los genes y el entorno. La figura 12-3 muestra un gran descenso en la tasa de tuberculosis antes de que se estableciera el tratamiento eficaz en la década de 1950.

booksmedicos.org

12-4 **B.** En ausencia de sesgos, la significación estadística establece que es poco probable que una asociación sea debida al azar, mientras que los criterios de Bradford Hill van más allá para establecer si una asociación es causa-efecto.

12-5 **C.** Los análisis de costo-efectividad describen los costos financieros de los efectos clínicos, como los años de vida salvados, como en este ejemplo, sobre las pruebas y tratamientos alternativos.

12-6 **D.** Cuando la exposición y la enfermedad se miden en los grupos, no en los individuos, es posible que los individuos que presentaron la enfermedad no fueran los expuestos, y que las personas expuestas no fueran las que presentaron la enfermedad, un problema denominado «falacia ecológica».

12-7 **B.** En un estudio de series temporales, un cambio en la tasa de enfermedad después de la intervención podría estar causado por otros cambios en las condiciones locales en el mismo momento. Esta posibilidad debe ser descartada antes de tener la confianza de que la exposición del estudio causó el resultado.

12-8 **B.** Los estudios controlados aleatorizados, si se diseñan y se llevan a cabo correctamente, son la evidencia más sólida de causa y efecto. Pero no es posible realizarlos para todas las causas posibles porque no es ético involucrar a las personas en estudios de daño y también porque podrían no estar dispuestos a cooperar en un ensayo de este tipo.

12-9 **E.** El patrón de evidencia de todos los criterios de Bradford Hill es más poderoso que la evidencia para cualquiera de ellos.

12-10 **D.** El intervalo de confianza es tan amplio que es consistente con el daño o el beneficio y, por lo tanto, no establece si hay una asociación entre los teléfonos móviles y el cáncer cerebral, y mucho menos si los teléfonos móviles son una causa.

CAPÍTULO 13 RESUMEN DE LA EVIDENCIA

13-1 **E.** La principal ventaja de combinar (agrupar) los resultados del estudio en una revisión sistemática es tener un tamaño de muestra más grande y, como consecuencia, una estimación más estable y precisa del tamaño del efecto. Aunque se podría argumentar que los resultados son más generalizables porque provienen de más de una ocasión o lugar, la generalización no es una ventaja principal.

booksmedicos.org

13-2 **D.** Se puede buscar en MEDLINE usando términos de contenido y métodos, pero no incluye todas las revistas del mundo, las búsquedas pueden pasar por alto algunos de los artículos de MEDLINE y no son un remedio para el sesgo de publicación, ya que todas las citas se publican. Estas limitaciones explican por qué los revisores deben usar otras formas complementarias para buscar artículos además de MEDLINE.

13-3 **B.** Aunque los estudios individuales suelen tener un poder estadístico limitado en los subgrupos, la combinación de varios estudios puede superar este problema, siempre y cuando se combinen los pacientes, no los ensayos.

13-4 **A.** Los puntos B-E son elementos de PICO, componentes importantes de la pregunta de investigación específica. Las covariables son un aspecto crítico de la investigación, especialmente para los estudios observacionales y para identificar la modificación del efecto en los ensayos clínicos, pero tratan acerca de cómo se puede responder con éxito a la pregunta clínica y no sobre la pregunta en sí misma.

13-5 **B.** Las revisiones sistemáticas deben incluir una descripción de la validez metodológica de los estudios para ayudar a entender qué tan fuertes son las conclusiones. La combinación de los elementos individuales de calidad en una medida de resumen no está bien establecida porque no es adecuado darle el mismo peso a cada elemento.

13-6 **D.** La situación habitual es que los estudios negativos pequeños (aquellos que no encontraron un efecto) tienen menor probabilidad de ser publicados que los estudios pequeños que encuentran efectos. Es probable que los estudios más grandes se publiquen sin importar los resultados.

13-7 **B.** Las revisiones narrativas pueden adoptar un enfoque integral para establecer las preguntas que los médicos deben responder para tratar a un paciente, pero a expensas de una descripción transparente de la base científica de la evidencia citada.

13-8 **B.** Los diagramas de bosque resumen la evidencia cruda en una revisión sistemática. El tamaño del efecto combinado y el intervalo de confianza pueden o no estar incluidos, dependiendo de si es apropiado combinar los resultados del estudio.

13- **A.** No es posible controlar una covariable en un metaanálisis de

booksmedicos.org

9 estudios, como se podría hacer en los estudios principales o en un metaánalisis individual.

13-10 **A.** La combinación está justificada si hay relativamente poca heterogeneidad entre los estudios, según lo determinado en una revisión informada de los pacientes, las intervenciones y los resultados. Una prueba estadística para la heterogeneidad también es útil en general, pero no en este caso, con tan pocos estudios, debido a la baja potencia estadística.

13-11 **D.** Los modelos de efectos aleatorios tienen en cuenta la heterogeneidad, si no es extrema y, por lo tanto, los resultados de los intervalos de confianza son más amplios y tienen mayor probabilidad de ser más precisos que los intervalos de confianza más estrechos que resultarían de un modelo de efecto fijo.

CAPÍTULO 14 ADMINISTRACIÓN DEL CONOCIMIENTO

14-1 **C.** Es virtualmente imposible mantenerse al día con toda la investigación nueva científicamente sólida y clínicamente relevante en su campo sin la ayuda de una publicación que lo haga por usted.

14-2 **D.** Los periódicos son una fuente rica de información en muchos de los aspectos de ser médico, la historia, política, ciencia, ideas, experiencias y más, pero no son particularmente buenos para hallar respuestas a preguntas inmediatas o incluso vigilar los nuevos avances en su campo.

14-3 **D.** Los recursos enumerados de A-C son excelentes para buscar las respuestas a las preguntas clínicas, mientras que un artículo reciente, fuera de contexto con el resto sobre la misma pregunta, tiene un valor marginal a menos que sea mucho más sólido que todos los demás que lo precedieron.

14-4 **C.** Las búsquedas en PubMed son indispensables para determinar si se ha informado de un episodio infrecuente. Son una parte importante para encontrar todos los artículos publicados sobre una pregunta clínica específica y son útiles para los investigadores, pero son demasiado ineficaces para la mayoría de las preguntas en el lugar de la atención médica.

14- **B.** A pesar de que la revisión por pares y la edición hacen que los

booksmedicos.org

5 artículos sean mejores (más legibles, exactos y completos), los artículos están lejos de ser perfectos cuando el proceso ha terminado y son publicados.

14-6 **C.** El conflicto de intereses está relacionado con una actividad específica y no existe en el caso general de invertir en productos médicos que no se relacionan específicamente con la colonoscopia.

14-7 **A.** Los reactivos B-E son recursos valiosos para buscar las respuestas a las preguntas clínicas en el sitio de la atención médica (si están disponibles los ordenadores y los programas relevantes). Las revistas médicas tienen un gran valor por otras razones, pero no son útiles para este propósito.

14-8 **D.** Tanto los pacientes como los médicos bien podrían respetar a los expertos famosos, pero deben buscar una base más sólida –la organización que los respalda, los hechos y la fuente de estos hechos– cuando deciden si deben creerles.

14-9 **A.** La clasificación GRADE se creó para las recomendaciones sobre el tratamiento, pero no para otras preguntas clínicas.

14-10 **E.** Todos son elementos básicos de un plan completo para la administración del conocimiento, como se describió en este capítulo.

booksmedicos.org

Apéndice B

Lecturas adicionales

1. INTRODUCCIÓN

Epidemiología clínica

Feinstein AR. Why clinical epidemiology? Clin Res 1972;20: 821–825.

Feinstein AR. Clinical Epidemiology. The Architecture of Clinical Research. Philadelphia: WB Saunders; 1985.

Feinstein AR. Clinimetrics. New Haven, CT: Yale University Press; 1987.

Hulley SB, Cummings SR. Designing Clinical Research. An Epidemiologic Approach, 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2007.

Riegelman RIC. Studying and Study and Testing a Test, 5th ed. Philadelphia: Lippincott Williams & Wilkins; 2005.

Sackett DL. Clinical epidemiology. Am J Epidemiol 1969; 89:125–128.

Sackett DL, Haynes RB, Guyatt GH, et al. Clinical Epidemiology: A Basic Science for Clinical Medicine, 2nd ed. Boston: Little, Brown and Company; 1991.

Weiss NS. Clinical Epidemiology: The Study of the Outcomes of Illness, 3rd ed. New York: Oxford University Press; 2006.

Medicina basada en la evidencia

Guyatt G, Rennie D, Meade M, et al. User's Guide to the Medical Literature: Essentials of Evidence-Based Clinical Practice, 2nd ed. Chicago: American Medical Association Press; 2008.

Hill J, Bullock I, Alderson P. A summary of the methods that the National Clinical Guideline Centre uses to produce clinical guidelines for the National Institute for Health and Clinical Excellence. Ann Intern Med 2011:154:752–757.

Jenicek M, Hitchcock D. Evidence-Based Practice: Logic and Critical Thinking in Medicine. Chicago: American Medical Association Press; 2005.

Straus SE, Glasziou P, Richardson WS, et al. Evidence-Based Medicine: How to Practice and Teach It, 4th ed. New York: Elsevier; 2011.

Epidemiología

Friedman GD. Primer of Epidemiology, 5th ed. New York: Appleton and Lange; 2004.

Gordis L. Epidemiology, 4th ed. Philadelphia: Elsevier/Saunders; 2009.

Greenberg RS, Daniels SR, Flanders W, et al. Medical Epidemiology, 4th ed. New York: Lange Medical Books/McGraw Hill; 2005.

Hennekins CH, Buring JE. Epidemiology in Medicine. Boston: Little, Brown and Company; 1987.

Jekel JF, Elmore JG, Katz DL. Epidemiology, Biostatistics and Preventive Medicine, 3rd ed. Philadelphia: Elsevier/Saunders; 2007.

Rothman KJ. Epidemiology: An Introduction. New York: Oxford University Press; 2002.

Campos relacionados

booksmedicos.org

Brandt AM, Gardner M. Antagonism and accommodation: interpreting the relationship between public health and medicine in the United States during the 20th century. Am J Public Health 2000;90:707–715.
Kassirer JP, Kopelman RI. Learning Clinical Reasoning. Baltimore: Williams & Wilkins; 1991.
Sox, HC, Blatt MA, Higgins MC, et al. Medical Decision Making. Philadelphia, American College of Physicians, 2006.
White KL. Healing the Schism: Epidemiology, Medicine, and the Public's Health. New York: Springer-Verlag; 1991.

2. FRECUENCIA

Morgenstern H, Kleinbaum DG, Kupper LL. Measures of disease incidence used in epidemiologic research. Int J Epidemiol 1980;9:97–104.

3. ANORMALIDAD

Feinstein AR. Clinical Judgment. Baltimore: Williams & Wilkins; 1967.
Streiner DL, Norman GR. Health Measurement Scales—A Practical Guide to Their Development and Use, 3rd ed. New York: Oxford University Press; 2003.
Yudkin PL, Stratton IM. How to deal with regression to the mean in intervention studies. Lancet 1996;347:241–243.

4. RIESGO: PRINCIPIOS BÁSICOS

Diamond GA. What price perfection? Calibration and discrimination of clinical prediction models. J Clin Epidemiol 1992;45:85–89.
Steiner JF. Talking about treatment: the language of populations and the language of individuals. Ann Intern Med 1999;130:618–622.

5. RIESGO: EXPOSICIÓN A LA ENFERMEDAD

Samet JM, Munoz A. Evolution of the cohort study. Epidemiol Rev 1998;20:1–14.

6. RIESGO: DE LA ENFERMEDAD A LA EXPOSICIÓN

Grimes DA, Schulz KF. Compared to what? Finding controls for case-control studies. Lancet 2005;365:1429–1433.

7. PRONÓSTICO

Dekkers OM, Egger M, Altman DG, et al. Distinguishing case series from cohort studies. Ann Intern Med 2012;156:37–40.
Jenicek M. Clinical Case Reporting in Evidence-Based Medicine, 2nd Ed. New York: Oxford University Press; 2001.
Laupacis A, Sekar N, Stiell IG. Clinical prediction rules: a review and suggested modifications of methodologic standards. JAMA 1997;277:488–494.
Vandenbroucke JP. In defense of case reports. Ann Intern Med 2001;134:330–334.

8. DIAGNÓSTICO

booksmedicos.org

McGee S. Evidence-Based Physical Diagnosis. New York: Elsevier; 2007.
Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978;299:926–930.
Whiting P, Rutjes AWS, Reitsma JB, et al. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern Med 2004;140:189–202.

9. TRATAMIENTO

Friedman LM, Furberg CD, DeMets DL. Fundamentals of Clinical Trials, 3rd ed. New York: Springer-Verlag; 1998.
Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. Ann Intern Med 2006;145:62–69.
Pocock SJ. Clinical Trials: A Practical Approach. Chichester: Wiley; 1983.
Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. N Engl J Med 1979;301: 1410–1412.
The James Lind Library. http://www.jameslindlibrary.org
Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA 2004;291: 1624–1632.
Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? Stat Med 1984;3:409–420.

10. PREVENCIÓN

Goodman SN. Probability at the bedside: the knowing of chances or the chances of knowing? Ann Intern Med 1999;130:604–606.
Harris R, Sawaya GF, Moyer VA, et al. Reconsidering the criteria for evaluation proposed screening programs: reflections from 4 current and former members of the U.S. Preventive Services Task Force. Epidemiol Rev 2011;33:20–25.
Rose G. Sick individuals and sick populations. In J Epidemiol 30:427–432.
Wald NJ, Hackshawe C, Frost CD. When can a risk factor be used as a worthwhile screening test? BMJ 1999;319: 1562–1565.

11. AZAR

Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. Ann Intern Med 1993;118:201–210.
Goodman SN. Toward evidence-based statistics. 1: the *P* value fallacy. Ann Intern Med 1999;130:995–1004.
Goodman SN. Toward evidence-based statistics. 2: the Bayes factor. Ann Intern Med 1999;130:1005–1013.
Rothman KJ. A show of confidence. N Engl J Med 1978; 299:1362–1363.

12. CAUSA

Buck C. Popper's philosophy for epidemiologists. Int J Epidemiol 1975;4:159–168.
Chalmers AF. What Is This Thing Called Science?, 2nd ed. New York: University of Queensland Press; 1982.
Morganstern H. Ecologic studies in epidemiology: concepts, principles, and methods. Ann Rev Public Health 1995;16:61–81.

booksmedicos.org

13. RESUMEN DE LA EVIDENCIA

Goodman S, Dickersin K. Metabias: a challenge for comparative effectiveness research. Ann Intern Med 2011;155: 61–62.

Lau J, Ioannidis JPA, Schmid CH. Summing up the evidence: one answer is not always enough. Lancet 1998;351: 123–127.

Leeflang MMG, Deeks JJ, Gatsonis C, et al. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008;149:889–897.

Norris SL, Atkins D. Challenges in using nonrandomized studies in systematic reviews of treatment interventions. Ann Intern Med 2005;142:1112–1119.

Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ 2010;340:c221.

14. ADMINISTRACIÓN DEL CONOCIMIENTO

Cook DA, Dupras DM. A practical guide to developing effective Web-based learning. J Gen Intern Med 2004;19:698–707.

Shiffman RN, Shekelle P, Overhage JM, et al. Standardized reporting of clinical practice guidelines: a proposal from the conference on guideline standardization. Ann Intern Med 2003;139;493–498.

booksmedicos.org

Índice alfabético de materias

Los números de página en cursivas indican figuras y los seguidos por una t indican tablas.

A

Abandonos, 103
Adherencia, 140
  al placebo, 162
Administración del conocimiento
  buscando respuestas a preguntas clínicas, 228-230
  definición, 225
  en práctica, 235
  guiando a pacientes en búsqueda de información, 234-235, 234t
  principios básicos, 225-228
  revistas, 231-234, 232t, 233, 234t
  vigilancia de nuevos avances, 230-231
Ajuste multivariado, 75
Aleatorización, 73, 139
  estratificada, 139
Análisis
  aclaratorio, 145
  de costo-beneficio, 205
  de decisión, 5, 205
  de Kaplan-Meier, 98
  de sensibilidad, 104-105
  de supervivencia, 95
  del mejor caso/peor caso, 104
  del tiempo transcurrido hasta acontecimiento, 99
  multivariado, 75
  por intención a tratar, 144
Analogía, 202

booksmedicos.org

Anomalía, 31
criterios, 41-45, *42*, *43*, *44*
distribuciones y, 38-41, *39*, 39t, *40*
variación y, 35-38, 35t, *36*, *37*, *38*
Años de vida ajustados por calidad (AVAC), 170
Aprendizaje (oportuno) justo a tiempo, 228
Asesoramiento, 157-158, *158*
Asignación aleatoria, 139
Asociaciones reversibles, 201, 201
Atención preventiva, 152
Azar, 10, 14, 16, 237
abordajes, 175-176
análisis por subgrupos, 187-189, *188*, 189t
comparaciones múltiples, 185-187, 186t
detección de episodios infrecuentes, 185, 185
efectos acumulativos, 10-11, 10
estimaciones puntuales, intervalos de confianza, 183-185
métodos multivariados, 189-190
prueba de hipótesis y, 176-180, *176*
razonamiento bayesiano, 191-192
sensibilidad y especificidad, 117, *117*

B

Bases de datos clínicas, 148
Biblioteca Cochrane, 230

C

Calibración, 56
Calidad de vida relacionada con salud, 142
Cambios en el estilo de vida, 153
Características iniciales, 139
Carcinoma de intervalo, 163
Casos
definición, 23-24

booksmedicos.org

selección estudios casos y controles, 83
Causa/s, 194
distantes, 51-52
evidencia
en contra y a favor, 199-202, 199t, *200*
indirecta para, 198-199, *198*, 199t
inmediatas, 51
modelado, 204-205, *205*
múltiple, 195-196, *196*
ponderación de evidencia y, 205-206, *206*
principios básicos, 195-198, 195t, *196*, *197*
proximidad, 196-198, *197*
única, 195, 195t
Cegamiento, 141-142, *141*
simple, 142
Censura, 100
Ciencias poblacionales, 4
Clasificación de información, 226
Cociente de daños, 101
Cohorte/s, 21, 29, 62, 62t, 238
falsas, 102
inicial, 96
método, 84
Cointervenciones, 141
Colegas clínicos, 228-229
Comorbilidad, 135
Comparaciones múltiples, 185-187, 186t
Conflicto de intereses, 226
Confusión, 8-10, 9, 14, 71-72, 237
confirmación, 72
control, 72-76
estrategia global para controlar la, 75-76
por indicación, 147
variables, 72

booksmedicos.org

Conjunto de pruebas, 102
Consistencia, 201
Constructos, 33
Control/es, 72
  hospital y comunidad, 84
  selección para estudios casos y controles, 83
Covariables, 6, 71
Criterios
  estándar, 109
  exclusión, 135
  inclusión, 135
Cumplimiento, 140-141
Curso clínico, 94
Curva/s
  de supervivencia, 98-100, 99
    interpretación, 100
  epidémica, 26, 26, 89, 90
  ROC *(receiver operator characteristic)*, 57, 114-115, 115

D

5 D, 2-3, 2t
Datos
  características, 181-182
  continuos, 32, 46, 238
  de intervalo, 32-33
  dicotómicos, 32
  discretos, 33
  intervalo, 32-33
  nominales, 32
  ordinales, 32, 46, 238
  simplificados para diagnóstico, 108-109
Denominadores, 18
Densidad de incidencia, 22, 29, 238
Diagnóstico

booksmedicos.org

precisión resultado, prueba en, 109-11, *109*
pruebas múltiples en, 125-129, *126*, 127t, 128t
razones de verosimilitud en, 122-125, 123t, *124*, *125*, 125t
sensibilidad y especificidad en, 111-117, *112*, 113t, *114*, *115*
simplificación de datos, 108-109
valor predictivo, 117-122, *118*
Diagrama
de bosque (de Forest), 215, 215
de embudo, 212
Diferencias biológicas, variación resultante, 36-37, *37*
Discriminación, 56
Diseño casos y controles, 65
Dispersión, 38
Distribución
de frecuencia, 38
descripción, 38, *39*, 39t
frecuencia, 38
gaussiana, 40-41
normal, 40-41, *40*
real, 39
sesgada, 39
Doble ciego, 142
Dos colas, 179
Duración enfermedad, 20

E

Efectividad comparativa, 134
Efecto
de etiquetado, 167
Hawthorne, 138
placebo, 138
EMBASE, 230
Encuestas, 21
Endémico, 26, 27, 30, 238

booksmedicos.org

Enfermedad
aumentando momio preprueba, 120-122
brote, 89-90
falta de estándares, objetivos para, 110-111
previa, 168
resultados, 26
Enmascaramiento, 141
Ensayo abierto, 142
Ensayo/s
aleatorizados, 156
controlados, 134, 135
alternativas para, 147-148
cegamiento, 141-142, 141
contra estudios observacionales, 148
diferencias que surgen por, 139-141
entrecruzamiento, 146
ética, 135
evaluación de resultados, 142-143, 143t
grupos de comparación, 138, 138
intervención, 136, 138
limitaciones, 147
muestreo, 135-136, 137
no inferioridad, 145
racimos, 145
variaciones, 145-146
en racimos, 145
clínicos, 134
pragmáticos, 136
cruzados, 146
de efectividad, 143
de eficacia, 143
de equivalencia, 145
de fase
I, 149

booksmedicos.org

II, 149
III, 149
de no inferioridad, 145
de práctica clínica, 136
de superioridad, 145
N = 1, 146-147
simples grandes, 136
Entrecruzamiento, 141
Epidemiología, 4. *V. también* Epidemiología clínica
clínica
azar, 10, *10*
definición, 3-4
número y probabilidad, 6
poblaciones y muestras, 6, *6*
principios básicos, 6-12
problemas clínicos y preguntas, 2, 2t
propósito, 3-4, 6
resultados sobre salud, 2-3, 2t
sesgo (error sistemático) en, 7-10, 7t
variables en, 6
Equilibrio, 135
Equipo de entrenamiento, 102
Error
tipo I (a), 176
tipo II (B), 176
Escala/s, 33
de Karnofsky, 35
Especificidad, 111, 112, 202
de herramienta de predicción del riesgo, 56-57
definición, 113
equilibrio entre, 113, 113t
establecer, 115-117
Espectro de pacientes, 116
Estadística

booksmedicos.org

de concordancia, 56
inferencial, 176
Estadísticamente significativo, 176, 177
Estadístico c, 56
Estado de salud, 142
Estándar de oro, 109
consecuencia de imperfección, 111
Estandarización, 75
Estimación, 176
puntual, 183
Estratificación, 74
del riesgo en programas de tamizaje, 58
del riesgo, 54
pronóstica, 102
Estudio/s
casos y controles, 81-83, *82*
anidado, 84
basado en población, 83
características, 81t
diseño, 83-87
medición de exposición, 85-87
selección de
casos, 83
controles, 83-85
de cohorte, 21, 62-63, *63*
históricos, 63, *64*
prospectivos, 63-64
retrospectivos/históricas, 63
uso bases de datos médicas, 64-65
ventajas desventajas, 65-67, 66t
de incidencia, 21, 62
de prevalencia, 21, 21, 30, 106, 238, 242
usos, 28
de series temporales, 202, 203

booksmedicos.org

múltiples, 203
del efecto del tratamiento, 134
ecológicos, 202
experimentales, 134
observacionales, 62, 71
contra aleatorizados, 148
de intervenciones, 147-148
sobre efectos del tratamiento, 156-157
pronósticos elementos, 95-97
transversales, 21
Eventos infrecuentes, detección, 185, 185
Exactitud, 33, 118
Exposición, 51

F

Factores
de confusión, 72
no medidos, 76
potenciales, 72
de riesgo, 51
causalidad y, 53-54
exposición común a, 52
para elegir tratamiento, 58
para establecer momio preprueba en prueba diagnóstica, 58
para predecir riesgo, 54
para prevenir enfermedad, 59
pronósticos
definición, 93
diferencias, 93-94, 94
identificación, 100-101, 101
Falacia ecológica, 202
Falsos
negativos, 109, 176-177
positivos, 109, 176-177

booksmedicos.org

Fiabilidad, 34, 35
Fracción
atribuible a población, 69
de muestreo, 25, 36
Fraude científico, 226
Frecuencia
conveniencia de palabras como sustitutos de números, 18
distribución de enfermedad, tiempo, lugar, persona, 25-28
estudios de prevalencia e incidencia, 21-23
estudios, 23-25
incidencia, 18, 18, 19t
prevalencia, 18-19, 18, 19t
relaciones entre prevalencia, incidencia, duración de enfermedad, 19-20
tasas comúnmente utilizadas, 20, 20t
usos de estudios de prevalencia, 28

G

Generalización, 11-12, 106, 242
Grupo expuesto, 62
Guías de práctica clínica, 229-230

H

Herramienta
de evaluación del riesgo, 54
de predicción
clínica, 54
de riesgo, 54
calibración, 56
discriminación, 56
estratificación del riesgo, 57
sensibilidad y especificidad, 56-57, 57
usos clínicos de, 58-59
Heterogeneidad, 217
Hipertensión de bata blanca, 8, 8

booksmedicos.org

Hipótesis, 132
nula, 178
Historia natural, 94

I

Incidencia, 18-19, 18
acumulativa, 21-22, 29, 238
características de, 19t
densidad, 22-23
en relación al tiempo, 19
relaciones entre, prevalencia, duración y, 19-20
Incidentaloma, 169
Independencia, supuesto, 129
Inferencia, 6
Informe de casos, 101
Inmunizaciones, 153
Interacción, 76
Interpretabilidad, 35
Intervalo de confianza, 183
Intervención, 134
Investigación de servicios de salud, 4-5

J

Jerarquía de diseños de investigación, 199, 199t

L

Lectura de revistas, 233-234, 233, 234t
Libros de texto electrónicos, 229
Lugar
de atención, 228, 228t
distribución de enfermedad por, 27, 27

M

Marcador, 53

booksmedicos.org

Margen de inferioridad, 145
Medicina
basada en la evidencia, 4, 4t
personalizada, 54, 55t
Medición/es
del efecto, 67
desempeño, 33-35, 35
variación resultante, 35-36, 35t
Medida cruda del efecto, 71
MEDLINE, 230
Metaanálisis, 216
a nivel del paciente, 217
acumulativo, 219-221, *220*
combinación de estudios en, 216-219, 218
en red, 219
estudios observacionales y diagnósticos, 221
fortalezas y debilidades, 221-222, 222
Metarregresión, 219
Método/s
de incidencia, 164
numérico, 17
multivariados, 189-190
Migración de la etapa, 96
Modelado multivariado, 189
Modelo/s de efecto/s
aleatorios, 218
fijo, 217
matemáticos, 189
Modificación del efecto, 76-77
Momio/s, 122
números y, 6
posterior (posprueba), 118, 122
previo (preprueba), 118, 119, 122
Muestra/s, 6, 6

booksmedicos.org

aleatorias, 25
Muestreo
por conveniencia, 25, 29, 238
probabilístico, 25, 29, 238

N

Numerador, 18
Números
probabilidad y, 6

O

Ocultamiento de la asignación, 141

P

Pandemia, 26
Pareamiento, 74, 85
en sombrilla, 85
Período de
latencia, 51, 80
preinclusión, 141
prevalencia, 18, 30, 238
Persona, distribución de enfermedad por, 27-28
PICO, 210
Placebo, 138
Plausibilidad biológica, 201-202
Población/es, 6
definida, 25
dinámica, 22, *22*, 30, 238
en riesgo, 25
muestra de estudio, 25
Postulados de Koch, 195t
Potencia estadística, 181, 184-185
Precisión, 34
estadística, 183

booksmedicos.org

Preguntas
buscando las respuestas, 228-230
clínicas, 2, 2t
Prevalencia, 18, 18, 118
características de, 19t
en relación con tiempo, 19
período, 18
puntual, 18
relaciones entre, incidencia, duración y, 19-20
estimación, 119
puntual, 18, 29, 238
Prevención
beneficios contra daños de, 169-172
carga del sufrimiento, 156
efectividad del tratamiento, 156-159, 158
en ámbito clínico, 152-153
enfoque científico a clínica, 155, 155t
ensayos aleatorizados en, 156
niveles, 153-155, 153
primaria, 153-154
tratamiento, 156-158
pruebas de tamizaje
consecuencias imprevistas, 166-169
rendimiento, 163-166
y tratamientos a lo largo del tiempo, 169
secundaria, 154
tratamientos, 158
terciaria, 154
Proceso de referencia para aumentar momio preprueba de enfermedad, 120-122
Pronóstico, 93
análisis de sensibilidad, 104-105
curso clínico, historia natural de enfermedad, 94
definición, 93

booksmedicos.org

descripción, 97-100, 97t, 98, 99
reglas de predicción clínica, 102, 102t
series de casos, 101-102
sesgo en estudios de cohorte, 102-104
Prueba/s
de hipótesis, 175
de ji cuadrado (χ 2), 178
de una cola, 179
diagnóstica, 108
en paralelo, 126-127, 127t
estadísticas, 176, 178-179, 178t
no paramétricas, 178
seriadas, 128, *128*
sensibles uso, 113
PubMed, 230
Punto de corte, 113

R

Rango, 34
Razón de momios, 87-88
ajustados, 202-204, *203*
cruda, 89
de prevalencia, 88
Razonamiento bayesiano, 190-191
Razones de verosimilitud
cálculo, 124-125, 125t
definición, 122
momio, 122
razón para su uso, 123-124, 123t
seriadas, 128-129
uso, 122-123
Red de causalidad, 195
Reglas
de predicción clínica, 102, *102t*, 127, *128t*

booksmedicos.org

de toma de decisiones diagnósticas, 127
Regresión
a media, 45-46
logística, 75
Relación dosis-respuesta, 200-201
Reproducibilidad, 34
Restricción, 73
Resultado/s
compuestos, 142
de salud, 2-3, 2t
evaluación, 142-143, 143t
falso positivo en prueba de tamizaje, 166
intermedios, 72
múltiples, 187, 189
en estudios pronósticos, 96-97, 97t
Resumen
de evidencia, 209
estructurado, 234, 234t
Reversión causal, 147
Revisión/es
por pares (colegas), 232
narrativas, 209
sistemática, 210-216, 210t, 212t
tradicionales, 209
Revistas, 231-234, 232t, 233, 234t
lectura, 233-234, 233, 234t
Riesgo
absoluto, 67-68, 67t, 68t
atribuible, 68, 68t
a población, 69
cociente, 67t, 68
control de, 72-76
variables extrañas, 88-89
de confusión, 71-72

booksmedicos.org

de efecto de etiquetado negativo, 167
de resultado falso positivo, 166-167
definición, 50
descripciones simples de, 71
diferencia, 67t, 68
estudios de, 61-67
observacionales y causa, 76
formas de expresar y comparar, 67-71
interpretación del atribuible, relativo, 68-69, 69t
modificación del efecto, 76-77
poblacional, 69-71
atribuible, 69
predicción, 54-56
proporcional de Cox, 75
razón de momios, 87-88, 87
sobrediagnóstico (seudoenfermedad) en tamizaje del cáncer, 167-169
reconocimiento, 51-54
relativo, 67t, 68, 68t
estimado, 88
tomando otras variables en consideración, 71

S

Seguridad, 157
Sensibilidad, 111, 112
de herramienta, predicción del riesgo, 56-57
definición, 113
equilibrio entre, 113, 113t
establecimiento, 115-117
(respuesta), 34-35
Series de casos, 101-102, 106, 242
Sesgo
de anticipación diagnóstica, 160-161, *161*, 160t
de confusión, 8-10, *9*
de cumplimiento, 161, 162

booksmedicos.org

de duración, 161, 161
de medición, 8, 14, 16, 104, 237
de migración, 103
de muestreo, *12*, 103
de publicación, 211
de selección, 7, 14, 16, 237
definición, 7
del recuerdo, 86
efectos acumulativos, 10-11, *10*
en estudios de cohortes, 102-104
en observación clínica, 7, 7t
por clasificación errónea no diferencial, 104
sensibilidad, especificidad, 116-117
Seudoenfermedad, 167-169
Sobrediagnóstico, riesgo en tamizaje del cáncer, 167-169
Sobrepareamiento, 85
Sobrerrepresentación, 25
Subgrupos, 146
Supervivencia de una cohorte, 97-98, 99
Supuesto de independencia, 129

T

Tamaño de muestra, 181
Tamizaje, 153
aceptable para pacientes y médicos, 166
cálculo, 163-164, 164t
cambios, 169
de incidencia, 159
de prevalencia, 159
especificidad, 163
seguridad, 165-166
sensibilidad, 163
simplicidad, bajo costo, 164-165
valor predictivo positivo bajo, 164

booksmedicos.org

Tasa
de cumplimiento, 20, 20t
de mortalidad
de los casos, 20
infantil, 20, 20t
materna, 20t
perinatal, 20, 20t
de supervivencia, 20
Tendencia central, 38
Tiempo
cero en estudios pronósticos, 96
distribución de enfermedad por, 26-27, 26
-persona, 22
Toma de decisiones
compartidas, 12
cuantitativas, 5
médicas, 12
Tratamiento
adaptando resultados de ensayos a pacientes individuales, 146-147
alternativas para, 147-148
asignación, 139, 139t
cegamiento, 141-142, 141
contra estudios observacionales, 148
diferencias que surgen después, 139-141
ensayo
aleatorizados controlados, 134, 135
de efectividad en, 143-144, 143
de equivalencia, 144-145, 144
de no inferioridad, 145
de superioridad, 145
estudios, 134
observacionales de intervenciones, 147-148
ética, 135
evaluación de resultados, 142-143, 143t

booksmedicos.org

fases de ensayos clínicos, 148-149
grupos de comparación, 138, 138
habitual, 138
ideas y evidencia, 132-134
intención a tratar, 144-145, 144
intervención, 136, 138
limitaciones, 147
muestreo, 135-136, 137
variaciones, 145-146

V

Validación, 102
Validez, 33-34, 33t, 35
constructo, 34
contenido, 33
criterios, 33-34
de criterios, 33-34
del constructo, 34
del contenido, 33
externa, 11, *11*, 14, 237
interna, 11, *11*
Valor
de *P*, 177
predictivo
definiciones, 117-118
determinantes, 118-119
negativo, 118
positivo, 118
Variable/s, 6
confusoras, 72
dependiente, 6
extrañas, 6, 71
independiente, 6
Variación, 35

booksmedicos.org

aleatoria, 10
biológica, 36-37, 37
efectos, 37-38
mediciones, 35-36, 35t
resultado de
diferencias biológicas, 36-37, 37
medición, 35-36, 35t
total, 37, 38
Verdadero
negativo, 109
positivo, 109
Vigilancia, 155
poscomercialización, 149

booksmedicos.org